Multi-Parameter Linear Least-Squares Fitting to Poisson Data One Count at a Time

Wm. A. Wheaton, Alfred L. Dunklee[1], Allan S. Jacobson, James C. Ling, William A. Mahoney, and Robert G. Radocinski

Jet Propulsion Laboratory, California Institute of Technology, MS 169-327, 4800 Oak Grove Drive, Pasadena CA 91109

Submitted to *The Astrophysical Journal*
April 30, 1993

High Energy Astrophysics Group
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109

[1] 4931 Alta Canyada, La Cañada, CA 91011

ABSTRACT

A standard problem in gamma-ray astronomy data analysis is the decomposition of a set of observed counts, described by Poisson statistics, according to a given multi-component linear model, with underlying physical count rates or fluxes which are to be estimated from the data. Despite its conceptual simplicity, the linear least-squares (LLSQ) method for solving this problem has generally been limited to situations in which the number n_i of counts in each bin i is not too small, conventionally more than 5-:10. It seems to be widely believed that the failure of the LLSQ method for small counts is due to the failure of the Poisson distribution to be even approximately normal for small numbers. The cause is more accurately the strong anti-correlation between the data and the weights w_i in the weighted LLSQ method when $\sqrt{n_i}$ instead of $\sqrt{\bar{n}_i}$ is used to approximate the uncertainties, σ_i, ill the data, where $\bar{n}_i = \mathrm{E}[n_i]$, the expected value of n_i. We show in an appendix that, avoiding this approximation, the correct equations for the Poisson LLSQ (PLLSQ) problem are actually identical to those for the maximum likelihood estimate using the exact Poisson distribution.

Since weighted linear least-squares involves a kind of weighted averaging, LLSQ estimators generally produce biased results when the data and their weights arc correlated. We describe a class of weighted 1)1,1, S() estimators which arc linear functions of the observed counts. Such PLLSQ estimators are unbiased independent of $\bar{n}_i$, even when the average number of counts in an entire *fit* is much less than one. Their variance is a minimum when the weights are calculated from the *true* variances of the data, but in general these are not accurately known. Fort-nately, the variance of the estimate is a very weak function of the weights near the optimum value, so for the PLLSQ problem it is easy in practice to find weights that arc virtually ideal, yet still completely unbiased. PLLSQ estimators which arc linear in the data also allow fitting multiple data sets by the calculation of only a scalar product, without the need to repeat the accumulation and solution of the LLSQ equations. Due also to the linearity of the estimates in the data, each count contributes to the answers independently of every other, so that the results for small bins are independent of the particular choice of binning. This property makes possible 1)1,1, S() methods which avoid binning the data altogether. Some alternatives to the approximation of the uncertainties in the data by the square root of the observed counts are discussed.

We apply tile method to solve a problem in high-resolution gamma-ray spectroscopy for the JPL High-Resolution Gamma-Ray Spectrometer flown on *HEAO* 3. Systematic error in subtracting the strong, highly variable background encountered in the low-energy gamma-ray region can be significantly reduced by closely pairing source and background data in short segments. Significant results can be built up by weighted averaging the net fluxes obtained from the subtraction of many individual source/background pairs. Extension of the approach to complex situations, with multiple cosmic sources and realistic background parameteriza-Lions, requires a means of efficiently fitting 10 data from single scans in the narrow (≈ 1.2 keV, for *HEAO 3*) energy channels of a Ge spectrometer, where the expected number of counts obtained per scan may be very low. Such an analysis system is discussed and compared to the method previously used.

Subject headings: Gamma-Rays: General, Numerical Methods

1. Introduction

Above 10 keV in high-energy astronomy, several observational problems become increasingly important: both source fluxes and detector effective areas usually decline, the background becomes more variable, and instrument telescope properties become less ideal. These facts make the data analysis for a hard x-ray or low-energy gamma-ray instrument, such as the Jet Propulsion Laboratory High-Resolution Gamma-Ray Spectrometer flown on the third High Energy Astronomy Observatory (*HEAO 3*), somewhat different from that at lower energy. We have developed a new analysis method for *HEAO 3* which has as its central objective the suppression of systematic errors in background subtraction. Accounts of the method, which we have called the "scan-by-scan" technique, appear in previous publications from our group (Riegler *et al.* 1981; Ling *et al.* 1983; Marscher *et al.* 1984; Mahoney *et al.* 1984; Ling *et al.* 1987; Wheaton *et al.* 1988). However, the weighted multi-parameter linear least-squares (LLSQ) fitting (Wheaton *et al.* 1983), which is essential to the success of the method in practice, has not been fully described previously although it has many advantages which make it of interest in its own right. Our main purpose here is to show how many of the traditional limitations of the LLSQ method for the analysis of Poisson data can be easily and completely overcome.

The reader should understand at the outset that linear least-squares fitting is not unique, despite the term "least"; nor even linear, as the method has often been implemented in the Poisson case ("PLLSQ" herein). There is instead a whole class of LLSQ fitting methods, distinguished by the means employed for weighting the data. Since for Poisson-distributed random variables the expected value is equal to the variance, the uncertainties in the data are never accurately known, nor are they usually the same for each datum. Thus some approximation to the uncertainties in the data is needed, which is equivalent to a choice of method for weighting the equations. For reasons which will be discussed, the obvious approximation, that of using each datum[1] n_i as the estimate of its own variance, cannot be recommended in general.

An important part of the scan-by-scan idea is to fit to short segments of data (for *HEAO 3*, one source scan, at most ≈ 20 min), so that the association of source and background is preserved. It follows that the fitting algorithm must work properly even when the expected number of counts in one scan is arbitrarily low, because for the low-background, high-resolution spectrometer on *HEAO 3*, the count rate in the narrow (≈ 1.2 keV) pulse height analyzer (PHA) channels drops from a few counts per hour to a few counts per week as the energy increases from 1 to 10 MeV (Figure 1; Wheaton *et al.* 1989). In addition to the requirement for validity at low counts, the method must be computationally efficient to allow the millions of independent fits needed—one for each energy channel (thousands), for each detector (four), and for each scan (thousands, in a typical observation). Finally, because of the need to perform many fits fully automatically, the method should not produce singular or almost-singular matrices except when they are unavoidable due to real linear dependence among the components of the model.

The most common methods of fitting linear models (models in which the expected data are linear functions of the unknowns) to Poisson data fail badly on one or more of these points. The method which is perhaps still the most common of all, the modified χ^2 method, which approximates the uncertainty in each datum n_i by $\sqrt{n_i}$ (*cf* section 4. and Eadie *et al.* 1971) fails on all of them. It requires a separate matrix inversion for each data set, is biased, and has a well-known tendency to yield singular matrices. Above all, the latter problems become worse for low counts, so that it is unusable for small numbers of counts per data bin. Most of these flaws can be traced to the approximation of the variance of the data in each bin i by the observed counts n_i rather the expectation value, $\bar{n}_i$. Even though $\bar{n}_i$ itself is not known to the experimenter, good alternatives to this approximation are available which essentially remove the difficulties noted above. While individual elements of our approach seem to be known to many practitioners (*e.g.*, Particle Data Group *et al.* 1990), the possibility of such a "one-count-at-a-time" method does not appear to be widely appreciated

In order to motivate the need for a better fitting method, and also to be as concrete as possible, we have closely tied the discussion to our experience with the *HEAO 3* scan-by-scan system, which has proved greatly superior to the more conventional superposition-style analysis (*cf* section 2.5.) used previously. This description of the context also serves to provide additional informa-

[1] Always taken to be the observed counts—never count rates—herein.

tion about the scan-by-scan analysis method as implemented for *HEA* O .3, describing its conceptual basis, and showing, using actual data, how it has given improved results, We have previously published (Wheaton *et al.* 1988) a study of the advantages of the scan-by-scan approach using Monte Carlo simulations of an idealized experiment. Some of tile benefits of tile scan-by-scan approach should be applicable to other experiments sharing similar analysis problems. Analysis of point sources by the Earth occultation method for BATSE (the Burst and Transient Source E-periment) on the *Compton* Gamma-Ray Observatory, has used a similar approach with very satisfactory results (Ling *et al.* 1993; Skelton c/ *al.* 1993).

In section 2. we discuss the *HEAO 3* context and the rationale for the scan-by-scan approach, with particular emphasis on tile problems presented by the background variability characteristic of experiments above shout 10 keV. In section 3. we establish some notational conventions and describe source and count rate models that have been used for *HEAO 3. In* section 4. we review the standard approach to weighted multi-parameter linear least squares fitting to Poisson data, **and show** its relation to weighted averaging, a relationship which clarifies the reasons for the problems often encountered. Section 5. shows that there are many simple and satisfactory alternatives to the approximatiouof the uncertainties in the data by the square root of the observed counts. Section 5. also gives a simple demonstration that, for the larger class of LLSQ methods noted in paragraph 2 of this section, linearity of the estimated answers in the data[2] implies complete unbiasedness in the low-count limit, independent of the statistical distribution of the data. In section 6. wc discuss the statistical uncertainties in the the fitted fluxes and the weighting of scans to obtain final answers. We also describe two fairly general strategies of unbiased weighting, which should be of use for other experiments. headers only needing a method of PLLSQ fitting which does not suffer from the defects noted above may wish to skip section 2. and go to sect.iolls 4.-6., referring to section 3. as necessary for our notational conventions. High-energy astronomers affected by the kincfof systcmatice rrorstflesc ail-l) y-seal) method seeks to eliminate may be more interested in sections 2. and 3..

2. Analysis Approach

Here we describe the *HEAO 3* context for our estimation problem. In section 2.1. we discuss the experiment, in section 2.2. the general problem of systematic error of background subtraction, and in section 2.3. we describe the superposition method and its defects. In section 2.4. we introduce tile "scan-by-scan" alternative and give its rationale as a method for suppressing systematic error of background subtraction. Finally in section 2.5. we show two comparisons of the alternative approaches. The first of these is based on Monte Carlo analysis of an idealized experiment intended to be as simple as possible and yet capture the essential difference between the superposition and stall-l)y-sca]l analysis. The second example is taken from *HEAO 3* data on orbit.

2.1. Experiment

The JPL gamma-ray spectrometer on *HEAO 3* had four high-purity germanium detectors operating from about 45 keV to 10 MeV. The total volume of germanium was about 400 cm^3; the total effective area was about 75 cm^2 at 100 keV. The cryostat was surrounded by a 6.6 cm thick CsI shield in active anticoincidence with the germanium detectors. The detector fields of view were 30° (FWHM) at low energy, increasing above a few hundred keV. Each event was energy-analyzed into 8192 PHA channels about 1.2 keV wide, and time-tagged to about 100 μs. Telemetry capacity allowed a maximum of 15.6 events pcr second to be individually transmitted to Earth per detector, compared to the typical ol~-orbit background rate of 10 s^{-1}; the average total dead time fraction was about 25%. The Ge prime sensors operated in the .50(I km, 43.6° inclination *HEAO* 3 orbit from shortly after launch on 1979 September 20 until cryogen exhaustion on 1980 June 1. The spacecraft spin axis normally pointed at the Sun, causing the instrument, which looked radially outwards, to scan a great circle perpendicular to the Ecliptic with every 20 min spin. In six months a complete survey of the sky was obtained. Details about the instrument and its radiation environment appear in Mahoney *et al.* (1980), Mahoney, Ling, & Jacobson (1981), and Wheaton (/ *al.* (1989).

[2] The inverse of the forward linearity of the model noted above

2.2. Systematic Error of Background Subtraction

Figure 1 shows an accumulation of' the experiment background spectrum. Also shown are the total spectrum of the Crab nebula and pulsar, and the 1σ Poisson noise level for a typical observation. Even the strongest sources are barely 30% of the background, dropping to a few percent in the MeV region. The background is the sum of many different components, such as the diffuse cosmic flux through the instrument aperture, gamma-rays from tile spacecraft and the Earth's atmosphere which leak through tile shield, activation] of instrument components by cosmically trapped radiations, and neutron interactions. The background is also a strong function of the geomagnetic coordinates, orientation, and activation history of the spacecraft. For low-Earth orbits the geomagnetic variables cause changes on a characteristic time scale of roughly 15 min. The amplitude of variation in the continuum ranges from some tens of percent at low energy to a factor of over five near 10 MeV. Finally, the functional dependence (on, for example, orbit parameters, experiment aspect, space radiation environmental conditions, and irradiation history) of the background is so complex that it has been impractical to construct a global (*i.e.*, valid for days or weeks, say) model for it to the accuracy (<< 1'%0) needed to do a background subtraction that approaches the statistical sensitivity limit. Thus one is forced to measure local background data associated with the source observation, and use them to estimate the background under the source.

By systematic errors in background subtraction we mean unmodel ed, non-Poisson errors arising due to the subtraction of an incorrect background model. The background model may be as simple as a single constant, or an elaborate semi-empirical parametrization. These systematic errors often become the factor effectively limiting the sensitivity of the experiment. Furthermore, since the distribution of their magnitudes is not known theoretically, in contrast, to Poisson statistical errors, it is difficult to place secure confidence bounds 011 the values of experimental results.

2.3. Superposition Approach to Analysis

Data from scanning x-ray experiments have often been analyzed by accumulating counts n_i and live times t_i (live time = clock time - dead time) in azimuthal bins i for long enough 10 obtain reasonable statistics. In the simplest method, the flux from a source is derived by designating an azimuthal region around the source position as "source", and adjacent regions as '(background". The source and background count rates are then estimated (by $[\sum_i n_i/\sum_i t_i]$, summed over the regions) and subtracted to give the rate due to the source alone. A more elaborate approach is to fit the run of accumulated azimuthal data, for example by a least-squares algorithm, to the response expected from a point source at the given position, taking the instrument angular response, or aperture function, into account. This allows analysis of multiple-source regions and more complex models for the background (eg., quadratic in azimuth), but from the point of view of this paper introduces no essential change. In either case, the method may be characterized as "first accumulate the data, then subtract the background".

While this method has been effective at x-ray energies (below about 10 keV), the circumstances described previously may combine to cause serious systematic error of background subtraction at higher energies. Figure 2 shows such an azimuthal accumulation, of *HEAO 3* data into 6° bins around the strong 667/668 keV background lines. Because of' the $\approx$ 30° FWHM aperture response, a cosmic point source should appear in the plot as a roughly triangular bump With a full width of at least 10 bins. No significant bump is evident, but the picture is confused by the presence of other features, spikes (eg., at 1890), dips (243°) and especially edges (42°- 60°, 216°- 2280), many obviously highly significant, which appear 10 be impossibly narrow for an experiment with *HEAO 3*'s broad aperture response. Such features are common, not only in 111,' ,40 3, but in many other scanning experiments operating above about 10 keV. It is puzzling to understand how they arise, as the time histories of the count rates are commonly as smooth as one can expect from counting statistics. However it is clear that such features make it impossible to carry out meaningfully the analysis scheme described above since the presence of a strong edge in the analysis region, like those in Figure 2, could overwhelm the formal statistical uncertainties of a subtraction or fit.

Since, loosely speaking, tile sum of smooth functions must be smooth, we are led to look for the origin of' this problem in tile many necessary selection tests and checks which must be applied to the data. Besides occasional data gaps, tests are necessary to remove data transmission errors, parity errors, and data affected by Earth blockage, high charged particle rates, South Atlantic Anomaly (SAA)" passages, and high magnetic latitude. The tests are typically made and applied independently. As a result, 20 min spacecraft spins are rarely complete, but are typically interrupted 2–3 times by the operation of these essential checks. Since the background is not constant, every time a data selection threshold is passed, an edge is introduced into the azimuthal accumul-tion of the count rate. If the background rate is highly variable, so that the edges are large, the noise they introduce exceeds that due to counting statistics. Since there are only 60 bins in Figure '2, and 70 or so spins per day, in a 30-day accumulation we expect an average of about 100 such selections per bin. The superposition of many such edges accounts for the disconcerting jaggedness in Figure 2.

If the scans do not sample the background randomly, the situation is worse yet. For example, the *HEAO 3 spin* rate was maintained within certain limits by the spacecraft control system, but the phase (i. e., the spin azimuth) was uncontrolled. Because the spacecraft was subjected to periodic tidal torques associated with its orbital motion, its spin could become locked, by a kind of resonance, to a harmonic of the orbital frequency, and the instrument would repeatedly view the same point of the sky from only a few pointsof the orbit.

In summary, in the presence of strong background variability, the "source" and "background" regions may contain data in which the true detector count rate varied over a wide range, rather than having single, well-de(il]cd, values. Since the true count **rates** in the two regions are not constant, experimental averages **of** them may be dominated by the particular sample 01 background **conditions** whit]] happened to be included. It 1ol-1ows that their difference may fail to converge to the cosmic source flux.

The first line of defense of the superposition analysis method against systematic background subtraction error has been to carefully sole'ct the data so as to reject regions in which background variation is a problem. Unfortunately at high energy this variation is so pervasive that if one attempts to formulate such restrictive selection criteria only a small fraction (for *HEAO 3*, < 10%) of the data survives. By strict selection criteria one effectively trades systematic errors for counting statistics errors. Eventually the data are **so** severely restricted that the Poisson uncertainties grow larger than the systematic uncertainties. This approach is not very satisfactory because many of the data arc discarded and because reliably estimating the uncertainties in the results remains problematical.

Another possibility would be to accumulate only data from those scans which are complete. Evidently this would entail an even larger loss of data with partial scans as common as they are. Analysis of regions with multiple sources, spread over a considerable range of azimuth, would become difficult or impossible. Yet since the time histories are substantially smooth, given any ***single*** scan containing the source of interest, even with some gaps, we could analyze the data for it in such a way as to extract the source rate, while avoidiug the edge problem. Following to its logical conclusion the idea of basing the analysis on scans leads to the **Stall-l)y-stall** method.

2.4. Scan-by-Scan Method

To suppress systematic errors of background subtraction for *H EA O 3*, we have reversed the usual "accumulate, then subtract" sequence of analysis. By this method, the source flux and its uncertainty σ_l arc estimated separately for each scan 1. The final estimate is a weighted average over scans, with the weights $w_l = \sigma_l^{-2}$. The uncertainties for each scan are estimated assuming Poisson statistics. Because the scans are generally incomplete, as explained in section 2.3., the uncertainties vary widely from scan to scan. The method in effect subtracts background for each scan individually, before accumulating scans to obtain significant answers, so that the association of the source and background **data** for each scan is preserved until the background has been removed. That the background may vary widely among scans thus causes no harm. **Such** pairing of' data and control is standard in the biological and social sciences, where unknown and uncontrollable sources of variation arc common. Although the statistical significance of the data from each scan is typically negligible, good statistics are recovered by the accumulation of the thousands of scans. Furthermore, estimates from scans taken during low-background portions of the orbit have smaller uncertainties on that account, and their higher weight can be preserved in the final average, instead of being lost when the counts from many scans are simply summed together, as they are in the superposition method.

Because the method fits to stretches of data that are small, the statistical uncertainties for a single scan are typically much larger than the magnitude of the systematic errors. While probably no really quantitative treatment of systematic error is possible, we imagine the systematic errors of background subtraction to be composed of two parts, one of which is uncorrelated with scans. The uncorrelated part should introduce a random error (small compared to the error due to counting statistics) into each scan, its sign varying from scan to scan, which will tend to cancel over many scans. Then when the net source flux estimates from L scans (typically $L > 1000$) are averaged to obtain the final answers, both the magnitude of the Poisson uncertainty and of the uncorrelated part of the systematic error should decrease together by similar factors, of order $L^{1/2}$ so that, if we could neglect the other, correlated, part of the errors, the total systematic error would remain insignificant in the average of many scans.

There is no guarantee that the systematic errors are entirely uncorrelated among scans, and some effects (especially the spin-orbit locking noted in section 2.3.) should be not be so. However, since the background variation is largely due to geomagnetic, orbit-related effects, and the spacecraft spin is nominally uncorrelated with the orbit, we expect most of the systematic error to decrease as described. Based on our experience with the improved results obtained, this seems to be the case.

2.5. Comparisons

Figure 3 shows an example from a Monte Carlo simulation (details appear in Wheaton *et al.* 1988), of how the variability of the background can dominate the statistical variance in a superposition analysis. Each of the 100 trials represents a complete *HEAO 3* observation (typically 1000 scans in six months) of a constant source in the presence of a strong background with 30% RMS variability. Data selection effects operated randomly on both the source and background regions. The upper histogram, (a), shows the frequency distribution of results for the 100 observations, each analyzed by first accumulating counts and live time for source and background for the 1000 scans, dividing to obtain average source and background rates, and finally subtracting to obtain the net source rate. The bottom panel, (b), shows results from the same 100 data sets, each analyzed by first subtracting to estimate the net rate for each source/background pair for the 1000 scans (100,000 estimates in all), and then performing a weighted-average of the 1000 net rates to obtain the best estimate for the data set. While both histograms have nearly the same mean, the true uncertainty in the estimates is given by their *scatter*, *i.e.*, the observed RMS width of the histogram. This width should be the sum in quadrature of a term due to Poisson statistics and an additional term due to residual systematic effects. Systematic errors broaden the (a) histogram by a factor of ≈ 1.5 relative to the Poisson errors. If detected at all, this extra uncertainty would reduce the information content (*i.e.*, the statistical weight) of the results in the upper, superposition, panel by a factor of $1.5 \times 1.5 = 2.25$ compared to the result (b) in the lower, scan-by-scan panel.

But the experimenter has access to the result of just *one* experiment, not 100. With a superposition analysis, he or she could easily be misled. Thus, of the 100 estimates, the upper histogram has three results more than 3σ above the true mean, while the scan-by-scan histogram has none. (For comparison, a histogram of 100 samples from a normally-distributed variable would average 0.14 samples more than 3σ above the mean.) By constructing frequency histograms of flux estimates, the scan-by-scan analysis can detect broadening due to residual systematic errors, so that they can be included in the .ncertainties reported.

Figure 4 shows an example of the success of .he scan-by-scan analysis method in removing obvious systematic errors from *HEAO 3* data. The strong 667/668 keV background lines (Wheaton *et al.* 1989) built up towards equilibrium during the first month of the flight. The superposition spectrum of the Galactic center shown in the upper panel, (a), was obtained by analysis of accumulations like that in Figure 2; it shows a strong spurious line due to inexact subtraction of the 667/668 keV background line. In the other spectrum, (b), obtained with the scan-by-scan analysis system, the spurious feature has been removed.

3. Linear Models of Observations

A basic characteristic of the scan-by-scan analysis for *HEAO 3* is that the data are fitted to a specific linear model, defined by the user. The model may contain cosmic point sources at specified positions on the sky, diffuse Galactic sources, and terms for various components of the background. In this section we describe our approach, define our notation, and describe some of the models that have been used

The scan-by-scan method often requires estimates of flux to be made based on only a few observed counts in the scan, and it is well-known that the standard approach to linear least-st[l]ares fitting to Poisson data fails for small numbers of counts in each bin. While a variety of "non-linear[3] methods" have been investigated and discussed in the literature which have given satisfactory results for small numbers of counts (Nousek & Shue 1989; Li & Ma 1983; Jansson 1984), they generally require solution of non-linear algebraic equations which depend on the data in a complicated way. For a high-resolution spectrometer with thousands of PHA channels, computational labor makes such approaches very unattractive.

Loredo & Epstein (1989) review and discuss arguments for a linear approach to inversion of Poisson data. Gamma-ray instruments are, by and large, *linear* devices. Since events are essentially independent, and are processed one-at-a-time, the counts $\bar{n}_{A+B}$ expected in an instrument due to the superposition of two sources, A and B, will be the sum of the counts $\bar{n}_A$ due to .4 and $\bar{n}_B$ due to B. Despite some exceptions (*e.g.*, pulse pile-up and dead-tilnc effects) in practice the linearity approximation is excellent. Arrays of counts form abstract vectors which may be added, subtracted, and multiplied by scalars with the usual algebraic properties (see, for example, Stewart 1973, Chapter 1). A gamma-ray instrument then corresponds to a linear transformation which maps a space $\mathcal{J}$ of photon (or background) sources into a data space, Z, of counts. Thus a natural language for gamma-ray astronomy data analysis is linear algebra. An important complication is that the linearity is not exact for observed data, but holds only in the expectation sense, for ensemble averages, so that the pure linear algebra becomes entangled with Poisson statistics. The standard data analysis problem is to pass in the opposite direction, from the vectors of observed Counts, to estimates of the underlying sources which account for them. Since it is mathematically unavoidable that the inverse of a linear transformation[4] must itself be another linear transformation, we take a linear approach herein.

A strictly linear least-squares estimator is guaranteed (*cf* the Gauss-Markov Theorem, Eadie *et al.* 1971, p. 135–136 and Appendix l),) to be unbiased even for finite samples. The theoretical advantages of, for example, the maximum likelihood method, hold asymptotically, with little guidaucc as to when the ideal properties of the limit arc reached in practice (Eadie *et al.* 1971, p. 155-156). We have been led, therefore, to examine the problems in the linear least-squares analysis of Poisson data more carefully. The result is a method which retains the advantages of a linear approach, even in the few-count limit, and is highly satisfactory in other respects.

The scall-lJy-stall analysis concept was originally developed with a simplified program (Riegler *et al.* 1981; Ling *et al.* 1983; Marscher *et al.* 1984) which allowed only one point source, a constant background, and three user specified energy bands, running on a 1977-vintage S.E.I,. 32/55 computer. Following experience with the initial version, a more capable code was introduced in 1983 which allowed a mix of up to eight cosmic source or background components in the model and 16 energy channels (Mahoney et *al.* 1984), limited by the computer's memory. After a period of evolutionary development the program was completely rewritten in FORTRAN-77 without, functional change, except to increase the model component and energy channel limits to 12 and 64 respectively. The increased speed and especially the increased memory available in a modest modern workstation make the current version effectively fifty times faster than its 1983 ancestor.

3.1. Notation

For convenience we summarize our notational conventions here. Indices are indicated by lower-case Roman subscripts, i, j, k, l; their range is always from one to the corresponding u~)per-case Roman letter, I, J, K, L. The subscript i always labels data bins; due to the scanning motion of 1//.'./10 3, these correspond to time bins. For the 300 FWHM instrument on *HEAO 3* we have typically used ≈ 20 s bins, corresponding to about 6° at the nominal spin period of 20 min. The transit of a cosmic source through the geometric aperture required roughly 100 s. Count rate components, whether cosmic or background, we always label by j. Our convention about i and j means that, without danger of confusion, we use σ_i to denote the uncertainty in the datum observed in bin i and σ_j for the uncertainty in an estimate of the j-th component rate[5], because i always refers to the data, and j always refers to the model (*cf* eq. [2]

[3] In this context, by common usage, "non-linear" means "not based On least-squares", since the standard modified χ^2 estimate is actually non-linear in the data.

[4] Restricted as necessary to the domain where it is non-singular.

[5] Note however that taken out of context, e.g. σ_1, is ambiguous.

below). Energy channels we label with k, but as each channel is treated entirely independently, the k indices have been suppressed wherever possible, as have also scans, indexed by 1, and the index for the four detectors. We use $E[u]$ for the expectation value of the random variable u, which as used herein, refers to the limiting value of the average that would be obtained if a variable could be sampled under exactly the same conditions many times without changing any of the true underlying count rates. We use $V[u]$ for the variance, $\sigma^2 \equiv E[u^2] - E[u]^2$, of u, and $Cov[u, v]$ for the covariance of u and v, defined as

$$\mathrm{Cov}[\tilde{u}, \tilde{v}] \equiv \mathrm{E}[(\tilde{u} - \bar{u})(\tilde{v} - \bar{v})]. \tag{1}$$

Here tildes stress that $\tilde{u}$, and $\tilde{v}$ are random variables, and e.g., $\bar{u} = E[u]$. '1'0 emphasize that a quantity is a statistical estimate, we may add a caret; thus $\hat{\mu}$ is an estimate of the true mean p = $E[u]$ of u.

3.2. Models for Count Rate

Each scan is treated as an independent experiment. Each energy channel k is also treated independently, for each detector, and analyzed by a separate fit. All the models assume the observed counts are Poisson-distributed about the expected counts $\bar{n}_i$ in each bin i of the scan slid that the $\bar{n}_i$ are sums of contributions from J components, which are to be estimated:

$$\bar{n}_i = \mathrm{E}[n_i] = \sum_{j=1}^{J} t_i T_{ij} r_j, \tag{2}$$

where t_i is the live time in the bin, T_{ij} are known proportionality factors, and the r_j are the unknown, underlying contributions due to cosmic sources or background components. We regard the array of counts $\{n_i\}$ as an I-dimensional vector of data to be expanded in terms of its components with respect to the J basis vectors (the *model vectors*, given by $\vec{A}_j = \{t_i T_{ij}\}$), with unknown expansion co-efficients r_j, which we wish to determine. We call the r_j of the model count rate "components" because they are the components of the expected data vector $\{\bar{n}_i\}$ expressed in terms of the model basis.

The experiment maps the model space of the $\{r_j\}$ into the data space of the $\{n_i\}$. The *expected* counts, $\bar{n}_i$, are actually contained in a smaller, J'-dimensional linear sub-space of the data space, with $J' \leq J$. If the map is full-rank (equivalent to the condition that the model vectors, which form the columns of the *design matrix* $\mathbf{A}$, be linearly independent; *cf* Stewart 1973), then $J' = J$. Finally Poisson noise operating on the expected counts $\bar{n}_i$ smears the observed counts out into the full I-dimensional data space.

For a cosmic point source j, the coefficient T_{ij} (*cf* Figure 5) is the instrument aperture response function for the bin i, computed from the source position, the energy, and the spacecraft aspect, and normalized to unity on the instrument viewing axis. For such sources we write

$$r_j = \eta A_0 \left(\frac{dF_j}{dE}\right) \Delta E \tag{3}$$

for the source count rate on the detector axis, where A_0 is the geometrical area, $\eta = \eta(E)$ is the full-energy peak efficiency at energy E, ΔE is the energy channel width, and dF_j/dE is the source differential photon flux. Prelaunch calibration data give the instrument response as a function of angles and energy (Mahoney *et al.* 1980). The form (3) takes no account of the non-diagonal energy response of the detectors due to Compton scattering and pair production. Strictly speaking spectral inversion, background subtraction, and spatial deconvolution should all be done together. Where needed this correction has been performed approximately as a separate step in the analysis, following those described here.

If j refers 10 a background component, the interpretation of T and r depend on the particular background model adopted For example, the simplest model with two cosmic point sources and a constant background would be

$$n_i = t_i \left[A + T_{i2} r_2 + T_{i3} r_3\right]. \tag{4}$$

Here $T_{i1} = 1$, the background $r_1 \equiv A$, and r_2 and r_3 are the cosmic source rates.

In practice, background variation within a scan can be usefully modeled and its accompanying systematic error largely removed. For example we may take the background rate in a given energy channel as

$$\mathcal{R}_b = A + BU, \tag{5}$$

where A is an unknown parameter which is constant within a scan, but might vary (e. g., due to the build-up of long-lived radioactive species) on longer time scales, B is an unknown proportionality factor, and U is the germanium detector upper level discriminator (ULD) rate, its threshold set at 10 MeV. This threshold is much lower than the energy ($\sim$ GeV) of the primary cosmic rays which are the ultimate source of most 0[the background, yet higher than most long-lived radioactive decay energies, so U largely counts shower secondaries. Therefore it is a good monitor of the local radiation environment and its prompt effects. Then the model for the expected counts in hin i would become, for a single cosmic source with on-axis rate r,

$$\bar{n}_i = t_i[A + BU_i + T_i r], \qquad (6)$$

with unknowns A, B, and r to be estimated.

Figure 5 shows an example of flL'~10 β live times, aperture response functions, and ULD values for a scan with $I = 30$ bins and $J = 5$ parameters in the model, including background, three cosmic sources (Cygnus X–1, Cygnus X–3, and the Galactic center), and the germanium ULD. The response for the background, a constant = 1, is not shown. The aperture response functions were evaluated at an energy of 70 keV.

4. Linear Least Squares as Weighted Averaging

Eadie *et al.* (19'71; chapters 7 and 8) discuss the theory and practice of estimation. They describe three alternative methods for estimating underlying component rates from binned event data. What they call the "modified minimum χ^2 method" has been especially widely used; we shall sometimes call it "the standard method" The adjective "modified" refers to the use of the observed data n_i in place of the expected values $\bar{n}_i$ to approximate the variances σ_i^2 in the weighted LLSQ method. Extensive discussions appear in, for example, Bevington (1969) and Blackburn (1970). FORTRAN programs implementing it in various contexts, together with other options for the estimated uncertainties, are given in Bevington (1969). The method has been derived by starting from the principle of maximum likelihood, approximating the theoretical Poisson distribution of counts by tile limiting normal distribution (which becomes exact in the limit of large expected counts $\bar{n}$), and then writing down the appropriate likelihood function for the normal distribution. However the same answer results from the classical solution to the abstract mathematical problem of solving an overdetermined system of linear equations in least squares, as obtained in the 19th century by Gauss (Wilks 1962; Gauss 1809).

4.1. Classical Least Squares

Given a general system of I linear equations in J unknowns x_j, $I \geq J$,

$$y_i = \sum a_{ij} x_j \qquad (7)$$

for $i = 1, \ldots, 1$; or in matrix form,

$$\vec{y} = \mathbf{A}\vec{x}, \qquad (8)$$

Gauss formed J *so-called* "normal equations", the solution of which minimizes the sum of the squares of the residuals of the original overdetermined system. For the j'-th unknown we multiply equation (7) by $a_{ij'}$ and sum over i, obtaining

$$\begin{aligned} \sum_i a_{ij'} y_i &= \sum_i a_{ij'} \sum_j a_{ij} x_j \\ &= \sum_j \left(\sum_i a^T_{j'i} a_{ij} \right) x_j, \end{aligned} \qquad (9)$$

where $a^T_{ji} = a_{ij}$ denotes transpose. Note that we may think o1" this as multiplying each of equations (7) by the weighting number $a_{ij'}$, and then summing over all the i to obtain an equation which is a weighted sum of' the I equations. Repeating this operation of weighting and summing for each of the unknowns, $j' = 1, \ldots, J$ in turn, we obtain a set of J different weightings of the original set of 1, or in matrix notation

$$\mathbf{A}^T \vec{y} = \mathrm{A}^T \mathrm{A} \mathrm{T}, \qquad (1o)$$

where again $\mathbf{A}^T$ is the transpose of the matrix A. The normal equations (9) form a J x J linear system, the normal matrix $\mathbf{B} = \mathrm{A}^T\mathrm{A}$ being square, and nonsingular if the columns of A arc linearly independent (Wilk s 1962). Inversion of II then yields values x_j which minimize the mean-square residual of the original overdctcrmined system (7).

In the context of Icast-squares estimation, each of equations (7) is the result ofa physical measurement of a quantity y_i. We wish to estimate the x_j. Since the above solution minimizes the sum of the squares of the residuals when the y: are simply regarded as abstract numbers, arbitrarily given, then if instead the y_i are random variables, experimental approximations to some underlying model (7), the classical solution gives the best-estimate x_j, in the sense 01' producing the best mean-square agreement with the data, without being based in any way upon the *distribution* of the y_i. The distribution of the y_i is nowhere used in the argument, being irrelevant to the problem of simply solving a system of linear equations in least-squares.

If the root-mean-square errors, σ_i, of the y_i are unequal but known, the method generalizes by multiplying each of equations (7) through by a weight, say $w_i^{1/2}$, and then carrying out the rest of the solution as before. Multiplication of each equation by $w_i^{1/2}$ is equivalent to left-multiplication of the matrix form (8) by a diagonal matrix W, whose elements are the $w_i^{1/2}$:

$$\mathbf{W}\vec{y} = \mathbf{W}\mathbf{A}\vec{x}. \qquad (11)$$

The application of the Gaussian prescription, with WA instead of A, then yields, (recalling that for any matrices A and W with a defined product $\mathbf{WA}, [\mathbf{WA}]^{\mathbf{T}} = \mathbf{A}^{\mathbf{T}}\mathbf{W}^{\mathbf{T}}$):

$$(\mathbf{A}^{\mathbf{T}}\mathbf{W}^{\mathbf{T}}\mathbf{W})\vec{y} = (\mathbf{A}^{\mathbf{T}}\mathbf{W}^{\mathbf{T}}\mathbf{W}\mathbf{A})\vec{x}, \qquad (12$$

with solution

$$\hat{\vec{x}} = [(\mathbf{A}^{\mathbf{T}}\mathbf{W}^{2}\mathbf{A})^{-1}\mathbf{A}^{\mathbf{T}}\mathbf{W}^{2}]\vec{y}, \qquad (13$$

since $W^T = W$. Note that, putting it all together, there are two successive steps of weighting: first by $w_i^{1/2}$, associated with the unequal uncertainties, and the second step pointed out in the discussion following equation (9), as an interpretation of the left multiplication of equation (8) by AT, resulting in equation (10).

The Gauss-Markov Theorem (discussed in standard texts, e.g. Graybill 1961; Eadie *et al.* 1971), shows that the weights, $w_i^{1/2}$ for each equation, should be $1/\sigma_i$ to obtain the optimal (minimum-variance) estimate of x_j. The two-step weighting, by which W appears only as W^2 in equation (13) then yields normal equations in which the weighting of each equation is finally by σ_i^{-2}, as we expect for a weighted average. It is a further consequence of the Gauss-Markov Theorem that the LLSQ estimators (13) are, with the above choice of weights, the unique minimum-variance, unbiased linear estimators, *independent of the distribution of the y_i and of the sample size*, if the expectation of the errors in the data y_i is zero and the errors are uncorrelated:

$$E[(y_i - \bar{y}_i)] = 0, \qquad (11)$$

and

$$E[(y_i - \bar{y}_i)(y_{i'} - \bar{y}_{i'})] = 0, \qquad (15)$$

for any distinct bins i and i', where $\bar{y}_i$ are the expected values of the data. Both of these conditions should always be satisfied for Poisson data. Because the solution for the x_j is obtained as a linear combination of the y_i, if in addition the errors in the y_i are normally-distributed, then the estimates for x_j will be also.

Note that the $\hat{x}_j$ are only linear in the y_i if the matrix W is independent of the y_i and x_j. One might at first think that for Poisson variables this *cannot* be the case! However, we are not required to assume W to be optimal, and we do not. While the minimum variance property of the theorem does require the particular choice $w_i = \sigma_i^{-2}$ above' for the weights, we show in Appendix D. that the unbiasedness depends only upon the linearity. Hence, we may instead just take W to be nearly optimal, but independent of the y_i and x_j. We return to this point in section 5..

In summary, if the data y_i are non-normal but have finite variance σ_i^2, the estimate (13) above yields the x_j which minimize the mean of the square of the normalized residuals for equations (i'). The solution is not dependent on the approximation of the Poisson distribution by a normal distribution, but the distribution of the estimates x_j will only be strictly normal if that of the y_i is also. It is shown in section 6. below that, for the Poisson problem, the weighted average of the results from many fits will be asymptotically normally distributed as the total number of counts, summed over fits (*i. e., scans,* in the case of *HEAO 3*) becomes large. But the crucial properties of unbiasedness and efficiency (see below) do not depend on the number of counts.

4.2. Application to Poisson Data

For Poisson data, where $\sigma_i^2 = \bar{n}_i$, the weighting matrix is given by $\mathbf{W}^2 = \mathrm{Diag}(1/\bar{n}_i)$ (the $I \times I$ diagonal matrix with diagonal elements $\bar{n}_i^{-1}$), and we obtain on writing out Equation (12),

$$\sum_i \frac{t_i T_{ij'} n_i}{\bar{n}_i} = \sum_j \left(\sum_i \frac{t_i^2 T_{ij'} T_{ij}}{\bar{n}_i} \right) r_j, \qquad (16)$$

with the correspondences $t_i T_{ij} \Leftrightarrow a_{ij}$, $n_i \Leftrightarrow y_i$, and $r_j \Leftrightarrow x_j$

These are the fundamental equations for the Poisson least-squares analysis problem. Since the expected counts $\bar{n}_i$ are functions of the r_j through the model equations (2), it is a non-linear system. Appendix A. shows that the same equations result from the exact application of the Principle of Maximum Likelihood to the Poisson distribution. As no approximation has been made above— beyond the assumption of the validity of the model equation (2) and of Poisson statistics—*both least squares and maximum likelihood lead to the same conclusion for the Poisson problem.* Furthermore, since equation (16) results from both, any solution derived from it must partake equally of the good theoretical properties of both the Principle of Maximum Likelihood and of least-squares estimation. Nevertheless, the insight that equation (16) is related to weighted averaging comes out of the least-squares approach, as does the recognition that it is "essentially linear", in a sense which will be made clear in section 5..

It seems almost irresistible to approximate the expected counts $\bar{n}_i$ by n_i, the observed counts, deriving from equation (16):

$$\sum_i t_i T_{ij'} = \sum_j \left(\sum_i \frac{t_i^2 T_{ij'} T_{ij}}{n_i} \right) r_j, \qquad (17)$$

which is linear in the r_j (but not in the n_i) and thus convenient to solve. This is the "modified minimum χ^2 method" described by Eadie *et al.* (1971). However this approximation has caused endless problems, despite having been used, at least at one time, nearly universally (Blackburn 1970, p. 52).

4.3. The Poisson Bias

These problems arise because each of the J normal equations, (9), is a weighted sum of the I data equations (7). But, if $\{u_i\}$ are random variables, drawn from populations which may be different for different i but have the common mean μ, then, the weighted average formula,

$$\hat{\mu} = \frac{\sum w_i u_i}{\sum w_i}, \qquad (18)$$

yields an unbiased estimate (i. e., $E[\hat{\mu}] = \mu$; see Eadie *et al.* 1971 for a discussion of bias and optimality of estimators) of $\mu = \mu_i = E[u_i]$ under very general conditions, independent of the distribution of the u_i and even *independent of the weights*—provided that the normalized weights w_i^* ($\equiv w_i / \sum w_i$) are uncorrelated with the data, u_i. A demonstration appears in Appendix B. The condition requiring no correlation will always be true if the w_i are not functions of any of the u_i, but not, in general, otherwise. The estimate is optimal (yields minimum variance of $\hat{\mu}$) when w_i^{-1} equals the variance of u_i.

The normal matrix $A^T W^2 A$ in the standard method, by equation (17), is a function of the n_i through the approximation $\sigma_i^2 \approx n_i$. It is clear that the resulting weights in the weighted average of the data equations (7) are strongly anti-correlated with the y_i, so that the system of equations (17) is biased. That is, every bin which has its observed counts lower than $\bar{n}_i$ receives too high a weight because its uncertainty is taken too low, and *vice versa*. These considerations account for the systematic underestimation (Particle Data Group *et al.* 1990; also noted by Bevington 1969, p. 248) encountered with the standard method.

The approximation $\sigma_i^2 \approx n_i$ is useless when $n_i << 1$. In such cases $n_i = 0$ (usually) or n_i = 1 (occasionally); whereas σ_i^2 always equals $\bar{n}_i$ exactly. The problems in equation (17) become spectacular when $n_i = 0$. Bevington (FCHISQ, p. 194) sets $\sigma_i^2 = 1$ in this case; again this has been common. If all the n_i are either 1 or 0, then the resulting w_i are all one, and the result is an unweighted fit. This is at least unbiased, although it will be far from optimal if—the only case in which weighting matters—the $\bar{n}_i$ vary widely, since the total variance of the averaging sum will then be dominated by those terms with the largest variance (*cf* eq. [18]).

The formation and solution of the normal equations has been partly superseded in modern numerical practice by singular value decomposition (SVD; see, *e.g.*, Press *et al.* 1986) using the QR algorithm (Stewart 1973). The reader may wonder if the validity of the argument above, based as it is on the normal equations, is affected when the normal equations are not used. The answer is no, because the solution for the best fit model is mathematically the same in either case. The application of SVD starting from the weighted least-squares equation (11) is possible regardless of whether the $\sigma_i \approx n_i$ approximation has been used to determine W. The results parallel those for the normal equations. The differences involve primarily effects of finite-precision arithmetic, especially in the treatment of nearly singular problems, which need not concern us here.

The bias due to the approximation $\sigma_i^2 \approx n_i$ is not limited to situations in which the expected counts are small, or to the classic multi-parameter linear least-squares method. It is instructive to apply the standard equation (17) to the simplest possible case, $J = 1$, obtaining a formula, which is completely wrong, for estimating a single count rate from binned data. This result is equivalent to a direct application of the weighted average formula (18) using the observed n_i to estimate the σ_i^2. Using instead $n = \sigma^2 = rt$ in equation (18), the unknown rate r cancels and we recover the correct formula, $\hat{r} = n/t = \sum n_i / \sum t_i$. Details appear in Appendix C..

Figure 6 shows the results of a simulation of a similar 1-parameter least squares situation, in which the bad effect of the $\sigma_i^2 \approx n_i$ approximation is clearly evident, as well as the virtually complete removal of the bias upon the substitution of a nearly uncorrelated estimate of $\bar{n}_i$ for n_i. It shows histograms of the estimates for 24,000 sets of simulated data for an x-ray source plus a known, constant background of $7.52\,s^{-1}$. The true (input) source rate on axis was $143.78\,s^{-1}$; the expected counts per bin ranged from 4.8 near the edge of the response function to about 30 at the peak. The upper panel, a), shows the histogram using the $\sigma_i^2 = n_i$ approximation. The unweighted mean of the histogram is $133.66\,s^{-1}$, its RMS width is $9.19\,s^{-1}$, and the bias for this many trials is $(133.66 - 143.78)/(9.19/\sqrt{24000})$, the magnitude of which exceeds 170σ. The lower panel, b), shows the histogram for the same data, but, after first computing the variance according to $\sigma_i^2 = n_i$, then iterating the calculation, using at each step the expected counts from the solution for the previous stage to obtain the current weights. The mean and RMS width, shown after two iterations (3 fits total) are now $(143.76 \pm 9.18)\,s^{-1}$, and the bias for the 24,000 trials is $-0.34\,\sigma$.

In the above case the bias is about $-1\,\sigma$ (*i.e.*, $[133.66 - 143.78]/9.19$) for each estimate, even though the $\bar{n}_i$ average about 15. The circumstance that the effect is typically marginally significant in a single fit, usually not bad enough to demand attention, is probably one reason it has been tolerated so long. If we calculate the bias due to the $\sigma_i^2 \approx n_i$ approximation analytically, we see that it is not even finite. No matter how large $\bar{n}$, given sufficient trials we will eventually encounter $n = 0$ in the infinite sum involved in the calculation of $E[\hat{r}]$. In the absence of some special action the expectation value of the estimate will be zero. Yet it will claim—on account of its infinite weight—to have zero error. With the substitution of one for all the zero data, the bias can be computed in terms of an exponential integral. Then it turns out that each data bin is biased about one count low, almost independent of $\bar{n}$, in the range $10 < \bar{n} < 100$.

The equations (17) have one further problem which the approach described below alleviates. It has often been noted that many-parameter least squares model inversions tend to encounter singular or very badly conditioned matrices. A partial cause of this difficulty is that putting random variables, that is, the observed counts, into the normal matrix may cause erratic problems, sometimes taking reasonably well-behaved matrices into hopelessly ill-conditioned or singular ones. It is a somewhat surprising fact that large matrices of random numbers often tend to be nearly singular. We understand the reason as follows. Since the determinant is geometrically the volume of the J-dimensional parallelepiped spanned by the columns of the matrix, if any of these column vectors is perturbed from its correct value (*i.e.*, as written in eq. [16]) into the plane of the others, that volume will be zero.

For the small number of parameters of *HEAO 3* analysis, ill-conditioned or singular matrices have not occurred except in circumstances where it was physically obvious that they were inevitable. Such situations arise in fitting sources very near to each other, or at nearly the same azimuth in the *HEAO 3* scan plane. More recent experience with larger problems (up to 1000 equations in 100 unknowns), supports this conclusion, that poor conditioning arises only when it is mathematically or physically inevitable, so long as observed counts do not appear in the design matrix.

5. Alternative Least-Squares Weightings

Because the scan-by-scan method often requires estimates of flux to be made based on only a few observed counts in the scan, the problems discussed in section 4. are intolerable. We have therefore recast the Poisson linear least-squares fitting algorithm in the light of the following considerations. To obtain an exact solution to the system, it would be sufficient to know the expectation (*i.e.*, exact) value of *any* independent subset of J of the I data equations (7). We would obtain this by averaging if we could repeat the entire experiment an infinite number of times. As this cannot be done, we are forced instead to use the normal equations, which are the best set of J linear equations that we can construct from the single set of I equations available. We obtain them as J different weighted sums of the data equations. But because of the unbiased property of weighted averaging, if $\mathrm{Cov}[w_i^*, u_i] = 0$ (*cf* eq. [B12] in Appendix B.), the expectation values of the equations (16) cannot depend on the weights used. And because of the minimum variance property of the $1/\sigma_i^2$ weighting, the variance of the estimate can only be a weak function (a quadratic form near its minimum) of the l-vector of weights near the extremum. Finally, multiplication of all the weights by the same factor does not change the weighted average.

5.1. Unbiased Weighting

These ideas hold the key to our problem. We can use our model for the physics, equation (2), directly with only approximate rates, but estimated independently of the n_i, to obtain rough $\bar{n}_i$ values for equation (16). In doing this, we are secure in the knowledge that not only is there no danger of wrong (i. e., biased) answers, but also that in the vicinity of the true $\bar{n}_i$ values, we can treat equations (16) as if their solution were almost independent of whatever estimates for the $\bar{n}_i$ we use. Then we replace equations (17) by equations (16), but with constants used to approximate the $\bar{n}_i$ in the denominators. Equations (16) become truly linear, in both the unknowns r_j and the data n_i, and are immediately solved in the usual way laid out in section 4..

More explicitly, if we replace the $\bar{n}_i$ in our basic equations (16) with some approximate constants $\bar{n}_i^*$, getting

$$\sum_i \frac{t_i T_{ij'} n_i}{\bar{n}_i^*} = \sum_j \left(\sum_i \frac{t_i^2 T_{ij'} T_{ij}}{\bar{n}_i^*} \right) r_j, \qquad (19)$$

and if $\hat{r}_j$ are the estimates of r_j found by solving these now linear equations (19), then the expected answers do not depend on $\bar{n}_i^*$:

$$\frac{\partial \mathrm{E}[\hat{r}_j]}{\partial \bar{n}_i^*} = 0, \qquad (20)$$

and, furthermore, near the true value $\bar{n}_i^* = \bar{n}_i$,

$$\left. \frac{\partial \mathrm{V}[\hat{r}_j]}{\partial \bar{n}_i^*} \right|_{\bar{n}_i^* \approx \bar{n}_i} \approx 0; \qquad (21)$$

that is, the efficiency of the estimate is only weakly dependent on $\bar{n}_i^*$. Equations (20) and (21) make more precise our earlier claim, that equations (10) are "essentially linear".

5.2. Insensitivity to Choice of Weights

It may at first appear that our procedure is impossibly circular, as information about the unknown answers must be assumed, *i.e.*, the weights (which are equivalent to $\bar{n}_i$), before a solution can be obtained. The same criticism could also be made of unweighted averaging, which in effect assumes that all the $\bar{n}_i$ are equal. In fact, unweighted averaging often works very well, never giving *wrong* (i.e., biased) answers, and usually being surprisingly efficient. For much the same reason, the approximate weighting we advocate has all these good qualities of unweighted averaging, but in addition gives estimates which are nearly optimal. Since the expectation value of the weighted average is independent of the weights, we are free to use any information available to determine them so long as we avoid the one taboo: *we must not look upon a datum while picking its weight*. Weighting amounts to multiplying each of the data equations through by a constant. Just as the solution of a non-singular square system is unaffected by this operation, the expected solution of an overdetermined system is also unchanged if any data equation (7), is multiplied by a weight. Hence, even a very poor estimate for the weights cannot introduce any bias into the results, it can only increase the variance of the estimate of the answers.

As expected, experience has confirmed that the variances of the results are very insensitive to the values used for the weighting. The results of experiments with real and simulated data (see below), done by reanalyzing the same data sets, while varying only the choice of weights, are completely in accord with the conclusion that any reasonable (accurate within a few tens of percent, say) values for the weights are adequate.

The reader may wonder why we bother to weight the data equations at all, since the values matter so little. Reflection about the weighted average formula, (18), indicates that if the variances of the data n_i are constant within about 10%, then the gain in efficiency of estimation due to weighting is quite small. If, on the other hand, the variances in the data vary by a factor of two or more, say, then in an unweighted average the terms in the sum with the largest variance can dominate the variance of the entire estimate, and the effect of the most accurate terms would be lost. In some experimental circumstances one might simply throw out the least accurate terms with negligible loss in efficiency, but in other situations, the weight of many terms, each with with large uncertainties, could equal or exceed the weight of a few terms with smaller uncertainties. Then the cost of discarding the many terms with larger uncertainties would be unacceptable. In deciding whether to weight or not one must consider both the dynamic range of the uncertainties and also their frequency distribution.

The need for accuracy in the weights parallels the importance of weighted versus unweighted averaging. For the case of *HEAO 3*, considering such effects as the (factor of several) geomagnetic variation in the rates at high energy, live time variations, different binnings, and data selection cuts, it is well worth the trouble to use weighted least squares when fitting to scan data. The scan illustrated in Figure 5, for example, contains two cases in which the normal $\approx 6°$ bins were split into unequal smaller pieces; the $\bar{n}_i$ and weights are affected correspondingly. The requirements on the accuracy arc, however, easily met. In other circumstances, where the count rate in a single fit shows great variation from bin to bin (for example, as in fitting to a spectrum with strong lines), weighting would be essential, and the need for an accurate estimate for the $\bar{n}$ would be somewhat more demanding. Yet still not severe: it is difficult to imagine a situation which would require better than 1OYO–2O% accuracy in $\bar{n}$ to avoid serious loss of sensitivity.

5.3. Implementation for Iii,'.to 3: The Relative Rate Vector $\vec{R}$

For *HEAO* 3, in the low energy gamma-ray region, the background is usually much the largest term in the total count rate, so $\bar{n}_i$ is roughly proportional to the live time t_i and just weighting each bin by $1/t_i$ is often sufficient. To allow for the possibility that other components than the constant term in the background may contribute significantly, wc define a vector $\vec{R}$ of rough "relative rates", supplied by the user, for each component of the model. These can be normalized by dividing each component by the background. The vector $\vec{R}$ is the incorporation of our strategic idea of using approximate *a priori* information, but independent of tbc observed data n_i, for the weighting.

Using this, we compute relative values for the expected counts in each bin, and thus obtain weights:

$$w_i^{1/2} = \left(\frac{1}{t_i \sum_j T_{ij} R_j}\right)^{1/2} \tag{22}$$

This choice weights each of equations (2) by a factor proportional (to the accuracy of our approximation for $\vec{R}$) to $1/\sigma_i$, so the weighted normal matrix $\mathbf{B} = \mathbf{A}^T W^2 A$ becomes

$$b_{jj'} = \sum_i w_i t_i^2 T_{ij} T_{ij'} \tag{23}$$

At the risk of belaboring the point, we stress again that the background information introduced via $\vec{R}$ is used *only* to weight the equations (which cannot change their expected solution), never to subtract the background. The latter would introduce severe systematic error if the background model were incorrect (for *HEA* O 3, by even 1%).

Table 1 shows the RMS errors obtained from a simulation in which the same Monte Carlo data sets were analyzed using seven different choices for $\vec{R}$, the first essentially exact and the others more or less incorrect. Wrong values for $\vec{R}$ produced only minor increases in the scatter of the estimated rates. Each row in the table summarizes the results for 10, ()D() different trials with simulated count data sets for the same scan. The same data sets were used for each row, and were analyzed identically except for the choice of the relative rate vector $\vec{R}$. The relative rates for the five components appear on the left, and the RMS scatter of the estimates obtained on tile right. The columns in the table are labeled by components j in the model. The live times and response vectors arc the same as in Figure 5. The "true" component count rates, used as input in generating the trial data, were: 6.0, 2.0, 0.2, 1.0, [s-l] and 0.03 [s^{-1} per ULD count s^{-1}], for the background, Cygnus X1, Cygnus X-3, the Galactic center, and the U 1,1) coefficient B, respectively.

5.4. Linearity of Solution in Observed Counts

Once the weights in equation (16) have been determined without recourse 10 the data, the solution becomes linear in the n_i so that the estimated rate r_j is

$$r_j = \sum_i \alpha_{ji} n_i, \tag{24}$$

where the α_{ji} are numbers, determined from the matrix inversion as functions of the t_i, the T_{ij}, and the weights, but not of the n_i:

$$\alpha_{ji} = t_i w_i^{1/2} \sum_{j'} \left[\left(\mathbf{B}^{-1}\right)_{jj'} T_{ij'}\right]. \tag{25}$$

Many advantages follow directly from the simplicity of equation ('2 1), in particular its linearity in the observed data n_i. It is very convenient computationally for the analysis of spectra, because it is not necessary to form and invert the normal matrix in every energy channel. All that is required is to generate the I-dimensional vectors $\vec{\alpha}_j$ (defined as $\{\alpha_{j1}, \alpha_{j2}, .., \alpha_{jI}\}$) once, and then form the scalar products $\vec{n}_k \cdot \vec{\alpha}_j$, where $\vec{n}_k$ is the I-vector of counts in energy channel k. Because the detector responses T_{ij} are slowly varying functions of energy, the normal matrix must in practice be formed and inverted at wide energy intervals, and the $\bar{\alpha}_j$ interpolated.

5.5, Validity at Low Count Rates

The folklore of the standard method requires that the number of counts per bin should not be too small (e. g., 5-30). in contrast, because the estimate, (24), is linear in the observed counts, there is now no difficulty no matter how small the expected counts $\bar{n}_i$ may be. in fact, it is easy to show that the expectation values of the estimated answers are exact independent of any size constraint on the n's:

$$\begin{aligned} \mathrm{E}[\hat{r}_j] &= \mathrm{E}\left[\sum \alpha_{ji} n_i\right] \\ &= \sum \alpha_{ji} \mathrm{E}[n_i] \\ &= \sum_i \alpha_{ji} \left(\sum_{j'} t_i T_{ij'} r_{j'} \right) \end{aligned} \tag{26}$$

and this will be identically equal to r_j if only the data in different bins are statistically independent and the condition

$$\delta_{jj'} = \sum_i \alpha_{ji} t_i T_{ij'} \tag{27}$$

holds, regardless of the distribution of the data. Here as usual $\delta_{jj'} = 1$ if $j = j'$ and zero otherwise. Relation (27) follows from the matrix inversion and the definition, (25), of the α_{ji} (*cf* Appendix 1).). Its validity does not depend on the choice, equation (2'2), for the weights, although of course tile variance of tile estimate does. The probability distributions of the final answers, being the averages of many such single-stall estimates, will be accurately normal by the Central Limit 'Theorem (*cf* section 6.) if only the total number of counts, summed over scans, is large, and will have the correct means and widths.

Figure 7 illustrates a set of Monte Carlo simulations of the scan in Figure 5. The live times t_i and response functions T_{ij} have been taken from data for the real scan. The rates r_j in the model were then chosen decreasing by successive factors of 100, so the counts per bin ranged from ≈ 100 (A), to ≈ 1 (B), to $\approx$ 0.01 (c). The expected counts were computed from the model equation (2) for each bin, and the observed counts found by Monte Carlo, using a Poisson random number generator. The resulting simulated scans were analyzed by the *HEAO 3* fitting subroutine. Finally, the results were tabulated for many identical scans, differing only in the Monte Carlo values for the observed counts. The number of scans for A and B were chosen to give the same number of total counts $N \equiv \sum_{i,l}(n_l)_i$, where $(n_l)_i$ is the number of counts in bin i of scan 1, in the simulated "experiment" of L scans. For C, L had to be reduced from the ideal of $L = 10^8$ due to computer time requirements and to the repeat cycle of the random number generator.

Figure 7 shows the results for Cygnus X-3. It is an especially demanding case, being a weaker source, only 8° (compared to the 30° FWHM of fl~,'.40 3) from a much stronger one (Cygnus X-1). The horizontal and vertical axes have been scaled by the appropriate theoretical factors (see Table 2) to keep the proportions of the histograms till, same.

While the distribution of the estimates for each scan ceases to be normal at the low rates in (C), where the expected count is about 0.3 per scan (30 bins x 0.01), yet the means and variances arc still correct The largest peak (off scale in the figure for the scale shown) at zero rate in (C) corresponds to that 70% of 'he scans with no counts at all. The next group of large peaks corresponds to scans with one count. The abscissa of each peak corresponds to the α_{ji} value for the particular bin in which the count occurred.

The corresponding histograms for the other four components are similar. The all have (*cf* Table 2) nearly the right width σ_t, and arc essentially centered at r, the true input rate used in the simulation. Here σ_t is the theoretical uncertainty in the estimation of r for a single scan, according to equation (28) in the next section. The actual observed centroid of the histogram of flux estimates for L scans is $\langle \hat{r} \rangle$, and its observed RMS width is (u). The difference, $(\langle \hat{r} \rangle - r) \equiv \Delta r$. The magnitude of Δr should be of order $\sigma_t/\sqrt{L} \equiv \sigma_L$. Finally, the significance rat io, I'/crr, , for each source detection depends on the total number N of counts in the experiment as $\sqrt{N}$, but not on the number of counts per bin.

Hence, paradoxical as it may seem, it is possible to do a background subtraction or a even a classical multiparameter linear least-squares fit to data which usually contain zero or at most one count, and obtain answers which are both correct and undegraded in the sense that the statistical uncertainties are what one would expect from the total number of counts in the overall data set.

One can verify by numerical computation (and it is straightforward to show analytically) that if the bins are not chosen too large, the α_{ji} are independent of bin size, so that the effect of a count does not depend on the binning. By equation 24, each count contributes an increment, α_{ji}, to r_j which depends on just the response T_{ij} at the generalized co-ordinates, q_i, of that count, independent of all other counts. It is this property which gives our paper its title. Thus a common objection to the use of binned data-that it throws away information due to the arbitrariness of the binning-is overcome, for the results obtained become independent of the bin size and boundaries once the size becomes small compared to the scale of variations in T_{ij}. It should even be possible to design data analysis systems which avoid binning altogether, by directly computing the effect of each event on $\hat{r}_j$ in terms of the response function, $T_j(q)$, at the point q_i where the event occurs. The computation would be performed event-by-event, rather than by bins, so that large arrays of empty data bins would not be needed. In such a method, the normal matrix would be computed in terms of scalar products of the response functions, by integration over the event co-ordinates q.

6. Estimation of Uncertainties

The standard method gives uncertainty and covariance estimates (e. g., Bevington 1969) from the elements of the inverse of the normal matrix, or covariance matrix. This is possible essentially because the normal matrix from equation (17), containing the observed counts as it does, has all the necessary information in it. The same information is present in equation (l(i), and it can be used in the same way if estimates of the $\bar{n}_i$ are available. However, if we have used only relative count rates to estimate the weights, as in practice we do for *HEAO* 3, an overall scale factor must be recovered. Appendix E. shows how the formulas below are related to the more familiar covariance matrix obtained from the standard method.

6.1. Basic Uncertainty Formulas

Uncertainties in the estimated rates can be found directly from equation (24) by

$$\sigma_j^2 = V[\hat{r}_j] = \sum_i \alpha_{ji}^2 V[n_i] = \sum_i \alpha_{ji}^2 \bar{n}_i, \qquad (28)$$

using the identity $V[a\tilde{u} + b\tilde{v}] \equiv a^2 V[\tilde{u}] + b^2 V[\tilde{v}]$ if a and b are constants and $\tilde{u}$ and $\tilde{v}$ are uncorrelated random variables. Since $\bar{n}_i$ is given by equation (2), then

$$\sigma_j^2 = \sum_{j'} C_{jj'} r_{j'}, \qquad (29)$$

where the $C_{jj'}$ are

$$C_{jj'} = \sum t_i T_{ij'} \alpha_{ji}^2, \qquad (30)$$

and are positive if $T_{ij'} \geq 0$. Equation (29) shows explicitly how the uncertainty in each unknown is produced by the true count rates in tbe problem. The uncertainty estimate (29) for $\hat{r}_j$ turns out to be the same as the uncertainty for a single parameter of interest obtained using the method of Avni (1975) in the linear case.

6.2. Application to Scan-by-Scan Analysis

The use of equation (29) to calculate the uncertainties requires some estimate of the r_j. It is possible 10 use the fitted answers from equation (24) directly, and for certain problems this may be the best or only available choice. however, in the overall *HEAO 3* scan-by-scan context, it is unsatisfactory for the following reasons. First, the estimates (24) of r_j for an individual scan may be negative. More importantly, we wish to use weighted averaging to combine estimates from the scans into final answers. The weights will be determined by the uncertainties in the r_j from equation (29), but if we used the data directly 10 estimate r_j, we would again have a situation in which the weights and data in the averaging of' equation (18) could be correlated. To avoid the possibility of introducing such a bias, we have estimated r_j in equation (29) independently. Minor generalizations of these methods suggest similar approaches which may be applicable to a variety of other experiments.

6.2.1. Uncertainties from Independent Local Data

One simple and robust solution, which works when the background is large and not too variable, is to estimate the background for the scan in the four *HEAO 3* detectors by summing the counts and live time for each. By using any three detectors we can estimate the background rate for the remaining one in an independent way. The same method is then applied to the other three detectors in turn. This method is adequate for *HEAO 3* below about 1 MeV, or in broad energy bands until geomagnetic background variability becomes large even witbin a scan, above about 2–3 MeV. It is convenient that the *HEAO 3* experiment consisted of four nearly identical detectors, but more generally even from a single counter one can always divide the data into several interleaved parts, and similarly estimate the weight for each part from the data from the others.

This method suggests that a particularly simple alternative solution to the problem of' determining uncorrelated weights for the LLSQ fitting would be to use, in equation (17), instead of n_i, an average of the data in neighboring bins. Even though a slight correlation remains (*cf* Appendix II.), such weighting should remove the bias to adequate accuracy for many situations. Problems associated with having data in the normal matrix - the need to re-invert, for each data set, and possible poor conditioning— would return, but this approach may still be the best method available for occasional use.

6.2..2. Uncertain ly-.. *from Rat c Modeling*

At high energy, tile previous ..3-detector" method may fail. First, the variability of the background even within a single scan becomes significant, because of the increasing amplitude of geomagnetic variation. Second, the count rate in a narrow channel becomes so low that there is often not even one count in the other three detectors. Then no estimate of the weight is available and the scan must be discarded. For such circumstances we have developed a method which fits the background in each energy channel to a model of the form (4): a constant plus a term proportional to the germanium detector ULD rate. The constants *A* and *B* are determined either from previous scans or from a table. This method has proved successful in the analysis of the 1809 keV ^{26}Al line (Mahoney *et al.* 1984). More generally) any model for the expected counts which has the accuracy needed for uncertainty estimation and which is unbiased, according to the principles in section 4., should be adaptable to this purpose.

6.3. Distribution of Flux Estimates

Because of tile linearity of the fitting and averaging over scans, the final flux estimates could in principle be written as a sum over all the observed counts, each with a constant co-efficient (*i.e.*, depending on the t's, T's, and true $\bar{n}$'s, but not depending on the data) determined implicitly by the procedures defined above. By the Central Limit Theorem (Eadie *et al.* 1971, Sec. 3.3.2) it follows that the estimated answers should themselves be nearly normally distributed about their true values if only the total number of counts in each energy channel, summed over all the scans in the observation, is not too small (by, say, the traditional 5–30 counts criterion). For *HEAO* 3, this is the case even for single PHA channels in the continuum at all energies (*cf* Figure 1) for typical source observations of 30 days (live time $t > 10^5$ s). On the other hand the tails of the distribution may differ substantially from those of a Gaussian for narrow-band effects observed at high energy in shorter times. The number of counts contributing to r_j must be borne in mind when translating sigmas into probabilities in such cases.

A sensitive test of the success of the uncertainty estimation and of the overall method is to make a histogram of standardized stall flux estimates about their mean. For each scan 1 we compute

$$z_l = \frac{u_l - \hat{\mu}}{\sigma_l}, \tag{31}$$

where u_l is the fitted estimate of the flux obtained for that scan, σ_l is the corresponding uncertainty, and μ is the estimated mean flux

$$\mu = \frac{\sum u_l w_l}{\sum w_l} \tag{32}$$

with $w_l = \sigma_l^{-2}$. The histogram of tile frequencies of occurrence of the z_l should be nearly normal, with zero mean and unit standard deviation, if the number of' counts in each scan >> 1. Even if the number of counts is small, the mean and RMS width of the histogram should still be zero and one, respectively. Figure 8 shows such a histogram, obtained from the analysis of a source from which no significant flux was observed (Marscher *et al.* 19s'4).

7. Discussion and Summary

In summary, our main results are as follows:

1. Systematic error in subtracting the strong, highly variable background encountered in the low-energy gamma-ray region can be significantly reduced by analyzing source and background data paired together in short segments. Significant results can be built up by the weighted averaging of many such segments.

2. Exact derivations for fitting Poisson data to linear models yield the same equations for the optimally weighted Linear Least Squares (LLSQ) and Maximum Likelihood methods, subject only to the correctness of the model. However, if the weights are not known exactly, a critical criterion which must be satisfied by least-squares algorithms is that the covariance of each datum and its corresponding normalized weight must be zero. This follows from the observation that classical linear least squares is an example of weighted averaging. According to the Gauss-Markov theorem, LLSQ statistical estimators which satisfy this criterion are virtually the best possible, in the following sense:

   (a) the results obtained are rigorously unbiased independent of the statistical distribution of the data, the number of samples, or other physically reasonable conditions on the weights;

   (b) with the optimal choice of weights, no other linear estimators give estimates with smaller RMS errors;

   (c) the above properties are true for small samples as well as asymptotically.

3. However, LLSQ estimators for Poisson data (PLLSQ estimators) as they have been implemented often fail to satisfy the covariance property in (2) above, typically due to the approximation of the variance by the observed counts; such estimators generally give biased results. The effect, being independent of the statistical distribution of the data, has nothing to do with the failure of the Poisson distribution to be approximately normal for small numbers of counts, contrary to a common belief. If zero data are replaced by ones, the $\sigma_i = \sqrt{n_i}$ approximation biases each data bin about one count low in the range of expected counts from 10 to 100.

4. Since the variance of the estimate is only a weak function of the vector of weights near the optimum (a quadratic near an extremum), it is not difficult to find PLLSQ weights which are essentially optimal for all practical purposes, so long as the critical property in (2) above is satisfied. Hence we can devise PLLSQ estimators which are rigorously unbiased and virtually optimal for arbitrarily low counts. We discuss various means for doing this in practice. For such estimators, in the limit as the bin size becomes small, the effect of each count becomes independent of the bin size or boundaries.

5. When analyzing successive data sets under constant experimental conditions using the type of PLLSQ estimators described in (4), the weights and the normal matrix remain constant. Then successive estimates reduce to the I-dimensional scalar product of the vector of observed counts with a *constant* vector computed from the inversion of the normal matrix, and the need for a matrix inversion for each data set is eliminated.

Many people have contributed to the *HEAO 3* scan-by-scan analysis system and to the ideas developed in this paper. W.A.W. is indebted for early discussions and insights to Duane Gruber, and especially to Frank Primini for explaining the reason for the bias in the usual modified χ^2 algorithm. Guenter Riegler and John Bradley made essential contributions to the *HEAO 3* analysis system. We also thank Richard Lamb, James Higdon, John Callas and especially Larry Varnell both for critically reviewing details of the method and for many helpful comments on previous drafts. Dorrene Comiskey patiently prepared numerous earlier versions of the manuscript. We thank the anonymous referee for carefully reading a difficult manuscript and suggesting important clarifications and improvements in the presentation. During the initial part of this work W.A.W. was supported by a grant from the National Research Council of the National Academy of Sciences. The work described in this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

Appendix

A. Maximum Likelihood Derivation of the Fundamental Equation

The basic equation (16), derived in the text by classical least-squares, also results from the application of the Principle of Maximum Likelihood to the Poisson distribution. A work on statistics.

such as Eadie *et al.* (1971), may be consulted for a general discussion of the maximum likelihood method. In this section we revert to the usual notation for the Poisson distribution function for the probability of observing n counts when the expected value of n is x:

$$P_n(x) = \left(\frac{x^n}{n!}\right) e^{-x} \tag{A1}$$

We also combine the live time t_i and instrument response T_{ij} into a single constant a_{ij} so that the model equation for the expected counts in bin i becomes (*cf* eq. [2] in the text):

$$x_i = \sum_{j=1}^{J} a_{ij} r_j . \tag{A2}$$

Then the likelihood L of obtaining the data set $\{n_i\}$ actually observed for a given set of J rate components $\{r_j\}$ is

$$L = \prod_{i=1}^{I} \left\{ \frac{x_i^{n_i}}{n_i!} e^{-x_i} \right\}, \tag{A3}$$

so that the log likelihood function $\mathcal{L} = \ln L$ is

$$\mathcal{L} = \sum_i n_i \ln x_i - \sum_i x_i - \sum_i \ln n_i! \tag{A4}$$

Setting the derivatives of $\mathcal{L}$ with respect to r_j to zero then gives the J likelihood equations:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial r_j} &= \frac{\partial}{\partial r_j} \sum_i n_i \ln \left[\sum_{j'} a_{ij'} r_{j'} \right] \\ &\qquad - \frac{\partial}{\partial r_j} \sum_{i,j'} a_{ij'} r_{j'} \\ &= 0. \end{aligned} \tag{A5}$$

Since

$$\frac{\partial}{\partial r_j} \left(\sum_{j'} a_{ij'} r_{j'} \right) = a_{ij}, \tag{A6}$$

this becomes

$$\begin{aligned} \sum_i n_i \frac{\partial}{\partial r_j} \ln \left[\sum_{j'} a_{ij'} r_{j'} \right] &= \sum_i a_{ij} \\ \sum_i \left\{ \frac{n_i}{\left[\sum_{j'} a_{ij'} r_{j'}\right]} \frac{\partial}{\partial r_j} \left[\sum_{j'} a_{ij'} r_{j'} \right] \right\} &= \sum_i a_{ij} \\ \sum_i \frac{n_i a_{ij}}{\left[\sum_{j'} a_{ij'} r_{j'}\right]} &= \sum_i a_{ij} . \end{aligned} \tag{A7}$$

Thus again we have a set of J non-linear equations in the J unknown rate's r_j. The denominator on the left is x_i in equation (A3). Thus we can write the following intuitive form:

$$\sum_i \left(\frac{n_i}{x_i}\right) a_{ij} = \sum_i a_{ij}, \tag{A8}$$

the factor in parentheses approaching 1.0 for large counts. This is an equation in the r_j through the dependence of x_i on r_j via the model equation (A2). Different as this seems from equation (16), if we multiply each term on the right by (x_i/x_i) and expand the numerator using the model we obtain:

$$\begin{aligned} \sum_i \left(\frac{n_i a_{ij}}{x_i}\right) &= \sum_i \left(\frac{a_{ij} x_i}{x_i}\right) \\ &= \sum_i \left(\frac{a_{ij} \sum_{j'} a_{ij'} r_{j'}}{x_i}\right) \\ &= \sum_{j'} \sum_i \left(\frac{a_{ij'} a_{ij}}{x_i}\right) r_{j'} \end{aligned} \tag{A9}$$

After substituting $t_i T_{ij}$ for a_{ij} and $\bar{n}_i$ for x_i, this turns out to be equivalent to equation (16) of the text:

$$\sum_i \left(\frac{t_i T_{ij} n_i}{\bar{n}_i}\right) = \sum_{j'} \sum_i \left(\frac{t_i T_{ij'} t_i T_{ij}}{\bar{n}_i}\right) r_{j'} \tag{A10}$$

B. Weighted Averages

If x_i are random variables drawn from I populations, $i = 1, \ldots, I$, all with the same mean μ, and finite but possibly different variances σ_i, if w_i are positive numbers or random variables, and if, for all i' and i, $w_{i'}$ and x_i are statistically independent, then the weighted average estimator $\hat{\mu}$ given by equation (18) is an unbiased estimate of μ, *i.e.*, $E[\hat{\mu}] = \mu$, independent of the distribution of the x_i or the w_i.

For, if we define $w_i^* \equiv w_i/(\sum w_i)$,

$$\begin{aligned} E[\hat{\mu}] &= E\left[\sum_i w_i^* x_i\right] \\ &= \sum_i E[w_i^* x_i] \\ &= E[w_i^*]\, E[x_i] + \mathrm{Cov}[w_i^*, x_i], \end{aligned} \tag{B11}$$

where the identity

$$E[\tilde{x}\, \tilde{y}] = E[\tilde{x}]\, E[\tilde{y}] + \mathrm{Cov}[\tilde{x}, \tilde{y}] \tag{B12}$$

for random variables $\tilde{x}$ and $\tilde{y}$, has been used in the last step. This follows from the definition of the covariance, equation (1) in the text. Because $\mathrm{Cov}[\tilde{x}, \tilde{y}] = 0$ if $\tilde{x}$ and $\tilde{y}$ are independent, and w_i^* is independent of x_i, the last term in the sum (B11) above is zero. Then

$$\begin{aligned} \mathrm{E}[\hat{\mu}] &= \sum_i \mu \mathrm{E}[w_i^*] \\ &= \mu \mathrm{E}\left[\sum_i w_i^*\right] \\ &= \mu \mathrm{E}[1] \\ &= \mu, \end{aligned} \qquad (B13)$$

since $\sum w_i^* = 1$. The standard choice for the weights, $w_i = 1/\sigma_i^2$, makes the variance of $\hat{\mu}$ a minimum.

The method of estimating the weights from "independent local data" described in section 6.21.01' the text, weakens the above condition in that it allows w_i to be a function of $x_{i'}$ for any $i' \neq i$. Then w_i is independent of x_i, but w_i^* is not quite, via the residual" effect of the presence of x_i in the normalizing sum. This appears to be small in practice.

C. One-Parameter case

We consider the problem of estimating a single count rate without any background at all, when the data have been binned into I bins, each with n_i counts observed in a time t_i. It is known that $n = \sum n_i$ and $t = \sum t_i$ are "sufficient statistics" (Lehmann 1959, pp 17–20) for this problem. That is, the maximally efficient estimator for the true rate r is a function of n and t only, so that the extra information due to the binning is s(Ipl'r[III-ous. Nevertheless it is interesting to compare algorithms for handling binned data in this simple situation. We may set $T_{ij} = 1$ without loss of generality.

Taking first the $J = 1$ case of equation (17) for the modified χ^2 method,

$$\sum t_i = \sum \left\{ \frac{t_i^2}{n_i} \right\} r, \qquad (C14)$$

and

$$\hat{r} = \frac{\sum t_i}{\sum (t_i^2/n_i)}. \qquad (C15)$$

As our second example, we consider the weighted average of the estimates for each bin, $r_i = n_i/t_i$, and estimate the measured uncertainties in each bin directly from the observed counts. Then $\sigma_i = r_i/\sqrt{n_i}$, and $w_i = t_i^2/n_i$. The weighted average formula gives

$$\begin{aligned} \hat{r} &= \frac{\sum w_i \hat{r}_i}{\sum w_i} \\ &= \frac{\sum t_i}{\sum (t_i^2/n_i)}, \end{aligned} \qquad (C16)$$

the same as that derived from equation (C15). Note that these results cannot be expressed in terms of $\sum n_i$ and $\sum t_i$ alone.

The third example is again weighted averaging, but recognizes that the true count rate r is the same, by hypothesis, in every bin. Thus the uncertainty should be derived from the expected counts $\bar{n}_i$ in each bin, with

$$\bar{n}_i = t_i r. \qquad (C17)$$

Then $\sigma_i = r/\sqrt{\bar{n}_i}$, $w_i = t_i/r$, and we obtain

$$\begin{aligned} \hat{r} &= \frac{\sum (n_i/r)}{\sum (t_i/r)}, \\ &= \frac{\sum n_i}{\sum t_i}, \end{aligned} \qquad (C18)$$

as the unknown r cancels.

This is clearly the right answer, and the one consistent with the known sufficiency of n and t, so we conclude that the answer (C15) and (C16) found by the other two methods is simply **wrong**. Since equation (17) in the text is plainly incorrect even for $J = 1$, its use for larger values of J seems difficult to justify.

D. The Gauss-Markov Theorem

Proofs and discussions of the Gauss-Markov Theorem appear in Eadie *et al.* (1971) and Graybill (1961). Here we only wish to show now the unbiasedness of the LLSQ linear estimators is independent of the distribution of the data and the sample size, but does demand that the design matrix and the weights be independent of the data.

Let the model equation be

$$\bar{\mathbf{n}} = \mathrm{E}[\vec{\mathbf{n}}] = \mathbf{A}\vec{\mathbf{r}} \qquad (D19)$$

where $\vec{\mathbf{n}}$ is the I-vector of observed counts, $\vec{\mathbf{r}}$ the true count rate J-vector, and A the $I \times J$ design matrix. Let the least-squares estimate of $\vec{\mathbf{r}}$ be

$$\hat{\mathbf{r}} = \alpha \vec{\mathbf{n}}, \qquad (D20)$$

where

$$\alpha \equiv (\mathbf{A}^T \mathbf{W}^2 \mathbf{A})^{-1} \mathbf{A}^T \mathbf{W}^2. \qquad (D21)$$

Let the weighting matrix W be some diagonal positive definite ($w_{ii} \geq 0$) matrix. Let both A and W be independent of $\vec{n}$.

Then

$$\begin{aligned} \mathrm{E}[\hat{\mathbf{r}}] &= \mathrm{E}[\alpha\vec{n}] & \text{(D22)} \\ &= \alpha\mathrm{E}[\vec{n}] & \text{(D23)} \\ &= \alpha\mathbf{A}\vec{\mathbf{r}} & \text{(D24)} \\ &= (\mathbf{A}^T\mathbf{W}^2\mathbf{A})^{-1}(\mathbf{A}^T\mathbf{W}^2\mathbf{A})\vec{\mathbf{r}} & \text{(D25)} \\ &= \mathbf{I_J}\vec{\mathbf{r}} & \text{(D26)} \\ &= \vec{\mathbf{r}}, & \text{(D27)} \end{aligned}$$

i.e., the expectation of the estimate equals the true value. Here I_J is the J x J identity matrix, on J. Note that this calculation does not require that n_i have any particular probability distribution so long as $\mathrm{E}[n_i]$ exists.

Tile critical step is from equation (D22) to equation (D23), where α has been treated as if it were a constant. This is based on the identity (B12) above for random variables. Thus in this context, "constant" (in some treatments the term "deterministic" has been used instead) means "independent of the n_i". The definition of α, equation (D21), shows the sufficiency of the condition that both A and W be independent of n_i. The weaker condition, replacing "independent of" with "uncorrelated with" could be substituted, and of course "nearly uncorrelated" may be adequate (*cf* Appendix II.) for the analysis of any real cxperimerrt.

E. Relation of Error Formulation to Covariance Matrix

The formulation for the uncertainties given in the text in section 6. was described for the case, usually applicable for *HEAO 3*, in which the background dominates the uncertainties as given in equation (29). Here we show how to correct this omission, and indicate also the relat ion to t he usual error formulation, in terms of t he covariance matrix.

Since terms other than A and B—for example, due to sources-could be significant in some situ - ations, we take approximate account of them by using tile relative rates R_j supplied by the user. We define tile normalization ξ to be the proportionality coefficient. between tile relative rate R_j and the true rate r_j:

$$r_j = \xi R_j. \tag{E28}$$

To estimate ξ we compare the observed rates with the $\vec{R}$ vector specified by the user. For example, if both the background and the UL D were explicitly present in the fit and the uncertainties were being obtained by the A + BU method, we would est imate

$$\xi = \frac{\sum t_i(A + BU_i)}{\sum t_i(R_A + R_B U_i)}. \tag{E29}$$

The sums are over bins i the scan, R_A and R_B are the A- and B-components of $\vec{R}$, and the A and B values are obtained independent of the current fit, as described in Section 6.2. .

Then from equation (28) the uncertainties follow by setting

$$\bar{n}_i \approx \xi t_i \left(\sum_j T_{ij} R_j \right), \tag{E30}$$

or in terms of the error coefficients $C_{jj'}$ as

$$\sigma_j^2 = \xi \left(\sum_{j'} C_{jj'} R_{j'} \right). \tag{E31}$$

The sum in equation (E31) turns out to be just $(\mathbf{B}^{-1})_{jj}$, the diagonal element of the inverse of the normal matrix. This corresponds to the usual relation $\sigma_j^2 = (\mathbf{B}^{-1})_{jj}$ of the standard formulation, in which $\xi =$ 1. The covariances of the answers can also be estimated from **equation** (E31) **in** the same way.

REFERENCES

Avni, Y. 1976, ApJ, 210, 642

Bevington, P. R. 1969, Data Reduction and Error Analysis for the Physical Sciences, (New York: McGraw Hill)

Blackburn, J. A. 1970, in Spectral Analysis, ed. J. A. Blackburn (New York: Marcel Dekker)

Eadie, W. T., Drijard, D., James, F. E., Roos, M., & Sadoulet, B. 1971, Statistical Methods in Experimental Physics, (Amsterdam: North-Holland), especially chapters 7 & 8

Gauss, K. F. 1809, *Theoria motus corporum coelestium in sectionibus conicis solem ambientium*, (Hamburg: Perthes & Besser)

Graybill, F. A. 1961, An Introduction to Linear Statistical Models, Vol. I (New York: McGraw-Hill), pp 114-116

Jansson, P. A. 1984, Deconvolution, with Applications in Spectroscopy, (Orlando: Academic Press)

Lehmann, E. L. 1959, Testing Statistical Hypotheses, (New York: John Wiley & Sons)

Li, T. & Ma, Y. 1983, ApJ, 272, 317

Ling, J. C., Mahoney, W. A., Wheaton, W. A., Jacobson, A. S., & Kaluzienski, L. 1983, ApJ, 275, 307

Ling, J. C., Mahoney, W. A., Wheaton, W. A., & Jacobson, A. S. 1987, ApJ, 321, L117

Ling, J. C., *et al.*, 1993, to be published in the Proceedings of the *Compton* Gamma-Ray Observatory Symposium, 1992 October 15–17, at Washington University, St. Louis, MO

Loredo, T. J., & Epstein, R. I. 1989, ApJ, 336, 896

Mahoney, W. A., Ling, J. C., Jacobson, A. S., & Tappborn, R. M. 1980, Nucl. Instr. & Meth., 178, 363

Mahoney, W. A., Ling, J. C., & Jacobson, A. S. 1981, Nucl. Instr. & Meth., 185, 449

Mahoney, W. A., Ling, J. C., Wheaton, W. A., & Jacobson, A. S. 1984, ApJ, 286, 578

Marscher, A. P., Brecher, K., Wheaton, W. A., Ling, J. C., Mahoney, W. A., & Jacobson, A. S. 1984, ApJ, 281, 566

Nousek, J. A., & Shue, D. R., 1989, ApJ, 342, 1207

Particle Data Group 1990 "Review of Particle Properties", Phys. Lett. B, 239, 1-516, Sec. III.28-38, "Probability, Statistics, and Monte Carlo"

Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. 1986, Numerical Recipes, (New York: Cambridge University Press)

Riegler, G. R., Ling, J. C., Mahoney, W. A., Wheaton, W. A., Willett, J. B., & Jacobson, A. S. 1981, ApJ, 248, L13

Skelton, R. T., *et al.*, 1993, to be published in The Proceedings of the *Compton* Gamma-Ray Observatory Symposium, 1992 October 15–17, at Washington University, St. Louis, MO

Stewart, G. W. 1973, Introduction to Matrix Computations, (New York: Academic Press)

Wheaton, W. A., Ling, J. C., Mahoney, W. A., & Jacobson, A. S. 1983, BAAS, 15, 940

Wheaton, W. A., Jacobson, A. S., Ling, J. C., & Mahoney, W. A. 1988, in Nuclear Spectroscopy of Astrophysical Sources, ed. N. Gehrels & G. H. Share, AIP Conference Proceedings No. 170, (New York: AIP), p. 511

Wheaton, W. A., Ling, J. C., Mahoney, W. A., Varnell, L. S., & Jacobson, A. S. 1989, in High-Energy Radiation Background in Space, ed. A. C. Rester, Jr. & J. I. Trombka, AIP Conference Proceedings No. 186, (New York: AIP), p. 304

Wilks, S. S. 1962, Mathematical Statistics, (New York: Wiley)

This 2-column preprint was prepared with the AAS LaTeX macros v3.0.

Table 1: Effect of Relative Rate Vector on Estimation Efficiency

Case	Relative Rate Vector Components:					RMS Scatter of Corresponding Estimated Rates:				
#	1	2	3	4	5	1	2	3	4	5
1	**1.0**	**0.3**	**0.03**	**0.2**	**0.0**	0.7856	1.680	2.829	3.152	0.04914
2	1.0	0.3	0.0	0.0	0.0	0.7857	1.690	2.829	3.152	0.04915
3	1.0	1.0	0.0	0.0	0.0	0.7858	1.683	2.835	3.154	0.04916
4	1.0	0.0	0.0	0.0	0.0	0.7857	1.681	2.831	3.152	0.04915
5	1.0	10.0	0.0	0.0	0.0	0.7900	1.804	3.062	3.190	0.04959
6	1.0	0.0	0.0	3.0	0.0	0.7891	1.682	2.832	3.177	0.04941
7	1.0	0.0	0.0	0.0	1.0	0.7948	1.687	2.838	3.171	0.04974

Table 2: Comparison Of Expected and Observed Estimates for ≈ 100, 1, and 0.01 counts-bin^{-1}

Case A. Approximately 100 counts per bin, L = 10,000 scans

j	Source	r	$\langle \hat{r} \rangle$	σ_t	$\langle \hat{\sigma} \rangle$	Δr	σ_L
1	Bkg	$6 \cdot 10^{0}$	$6.004 \cdot 10^{0}$	$7.86 \cdot 10^{-1}$	$7.89 \cdot 10^{-1}$	$3.7 \cdot 10^{-3}$	$7.9 \cdot 10^{-3}$
2	Cyg X-1	$2 \cdot 10^{0}$	$1.992 \cdot 10^{0}$	$1.68 \cdot 10^{0}$	$1.68 \cdot 10^{0}$	$-0.8 \cdot 10^{-2}$	$1.7 \cdot 10^{-2}$
3	Cyg X-3	$2 \cdot 10^{-1}$	$2.100 \cdot 10^{-1}$	$2.83 \cdot 10^{0}$	$2.84 \cdot 10^{0}$	$1.0 \cdot 10^{-2}$	$2.8 \cdot 10^{-2}$
4	G.C.	$1 \cdot 10^{0}$	$1.010 \cdot 10^{0}$	$3.15 \cdot 10^{0}$	$3.16 \cdot 10^{0}$	$9.6 \cdot 10^{-3}$	$3.2 \cdot 10^{-2}$
5	ULD	$3 \cdot 10^{-2}$	$2.979 \cdot 10^{-2}$	$4.91 \cdot 10^{-2}$	$4.93 \cdot 10^{-2}$	$-2.1 \cdot 10^{-4}$	$4.9 \cdot 10^{-4}$

Case B. Approximately 1 count per bin, L = 1,000,000 scans

j	source	r	$\langle \hat{r} \rangle$	σ_t	$\langle \hat{\sigma} \rangle$	Δr	σ_L
1	Bkg	$6 \cdot 10^{-2}$	$5.998 \cdot 10^{-2}$	$7.86 \cdot 10^{-2}$	$7.89 \cdot 10^{-2}$	$-2.2 \cdot 10^{-5}$	$7.9 \cdot 10^{-5}$
2	Cyg X-1	$2 \cdot 10^{-2}$	$1.995 \cdot 10^{-2}$	$1.68 \cdot 10^{-1}$	$1.68 \cdot 10^{-1}$	$-5.1 \cdot 10^{-5}$	$1.7 \cdot 10^{-4}$
3	Cyg X-3	$2 \cdot 10^{-3}$	$2.061 \cdot 10^{-3}$	$2.83 \cdot 10^{-1}$	$2.82 \cdot 10^{-1}$	$6.1 \cdot 10^{-5}$	$2.8 \cdot 10^{-4}$
4	G.C.	$1 \cdot 10^{-2}$	$9.867 \cdot 10^{-3}$	$3.15 \cdot 10^{-1}$	$3.11 \cdot 10^{-1}$	$-1.3 \cdot 10^{-4}$	$3.2 \cdot 10^{-4}$
5	ULD	$3 \cdot 10^{-4}$	$3.013 \cdot 10^{-4}$	$4.92 \cdot 10^{-3}$	$4.90 \cdot 10^{-3}$	$1.3 \cdot 10^{-6}$	$4.9 \cdot 10^{-6}$

Case C. Approximately 0.01 counts per bin, L = 1,000,000 scans

j	source	r	$\langle \hat{r} \rangle$	σ_t	$\langle \hat{\sigma} \rangle$	Δr	σ_L
1	Bkg	$6 \cdot 10^{-4}$	$5.97 \cdot 10^{-4}$	$7.86 \cdot 10^{-3}$	$7.85 \cdot 10^{-3}$	$-2.9 \cdot 10^{-6}$	$7.9 \cdot 10^{-6}$
2	Cyg X-1	$2 \cdot 10^{-4}$	$2.09 \cdot 10^{-4}$	$1.68 \cdot 10^{-2}$	$1.68 \cdot 10^{-2}$	$0.9 \cdot 10^{-5}$	$1.7 \cdot 10^{-5}$
3	Cyg X-3	$2 \cdot 10^{-5}$	$0.57 \cdot 10^{-5}$	$2.83 \cdot 10^{-2}$	$2.82 \cdot 10^{-2}$	$-1.4 \cdot 10^{-5}$	$2.8 \cdot 10^{-5}$
4	G.C.	$1 \cdot 10^{-4}$	$0.90 \cdot 10^{-4}$	$3.15 \cdot 10^{-2}$	$3.16 \cdot 10^{-2}$	$-1.1 \cdot 10^{-5}$	$3.2 \cdot 10^{-5}$
5	ULD	$3 \cdot 10^{-6}$	$3.12 \cdot 10^{-6}$	$4.92 \cdot 10^{-4}$	$4.93 \cdot 10^{-4}$	$1.2 \cdot 10^{-7}$	$4.9 \cdot 10^{-7}$

Fig. 1.— Strong source (Crab nebula) and background spectra (4-detector sum) for *HEAO 3*. The dashed line is the 1 σ noise level due to Poisson statistics, for a continuum observation of Cygnus X-1, a source in a favorable position for observation.

Fig. 2.— Two-week azimuthal data accumulation showing *HEAO 3* count rates versus scan angle for an energy band centered on the 667 and 668 keV background lines. Data away from the SAA, taken within 80° of the zenith, and with the McIlwain parameter $L < 1.6$, were used.

Fig. 3.— histograms of simulated count rate estimates obtained for 100 Monte Carlo trials of a typical *HEAO 3* observation consisting of 1000 source scans, in the presence of a strong background, when analyzed by first accumulating counts (top) and by first subtracting background (bottom). The *actual* uncertainties are the RMS widths of the histograms. The apparent errors arc obtained assuming Poisson statistics. The upper histogram is broadened by a factor of 1.5 with respect to both its claimed uncertainty and to the lower panel.

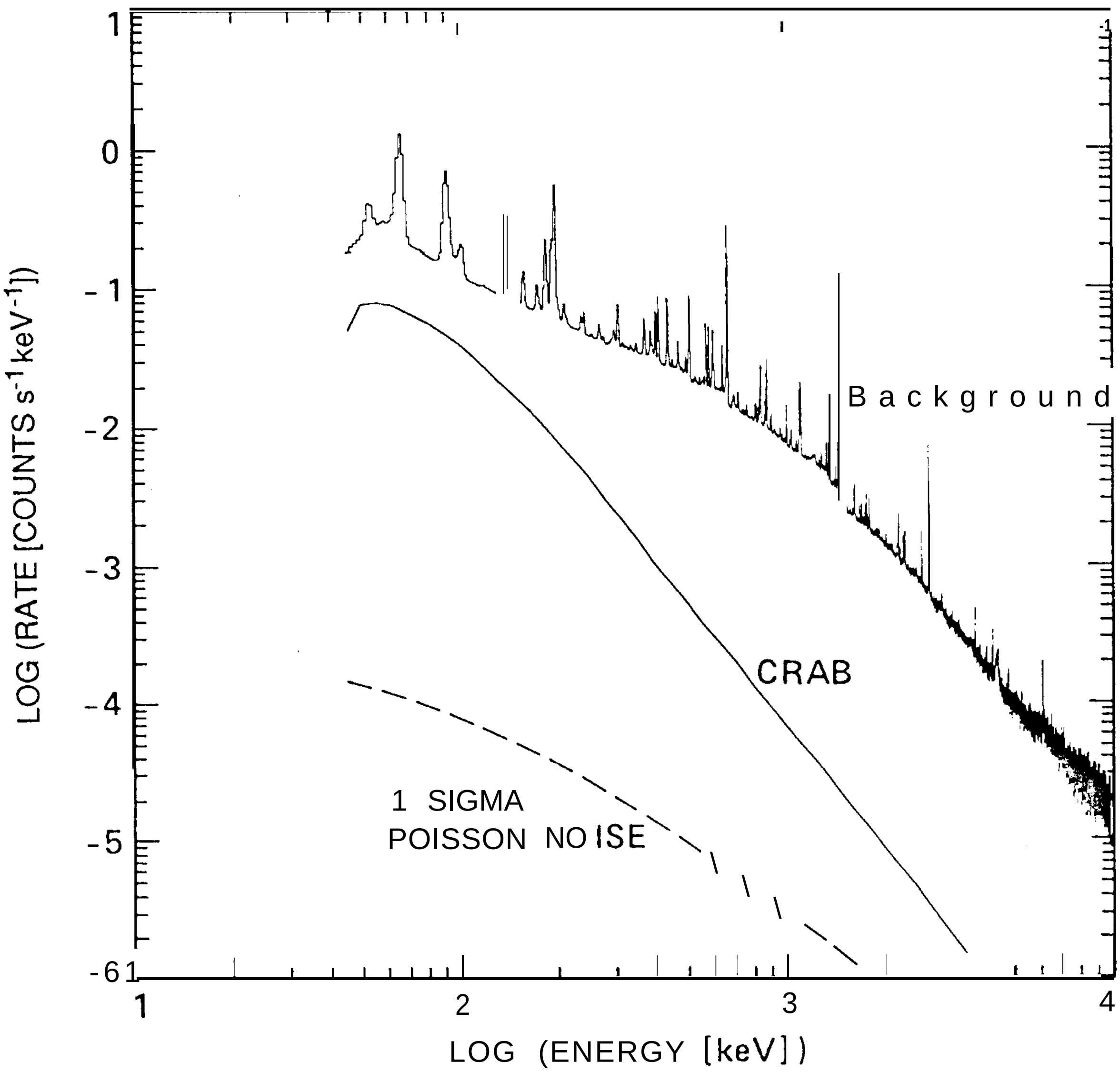
Fig. 4--- Comparison of *HEAO 3* analyses by the (a) superposition, and (1,) stall-by-stall methods, for the energy region containing the strong 667–668 keV background lines, for the Galactic center source (*cf* Figure 2). Note the elimination of the strong residual line feature, seen in (a), when the stall-by-stall analysis method is used.

Fig. 5.— Live time and response functions for a typical *HEAO 3* scan. a) Live time in bin, s; b) Aperture response for Cygnus X–1, normalized to 1.0 on axis; c) Same as b), for Cygnus X-3; d) Same as b), for the Galactic center; e) Germanium ULD rate s^{-1}.

Fig. 6.- Monte Carlo study showing the Poisson bias. The histograms are of rate estimates, for a simple 1-parameter problem described in the text, for the same 2 1,000 sets of simulated data, analyzed in two ways: (a) using the $n_i \approx \sigma_i^2$ approximation to estimate the uncertainties, and (b), after iterating the solution twice with the uncertainties based on the expected counts n_i from the model, rather than directly on the data, The means of the two histograms differ from the true value (143.78 count s^{-1}) by -170σ and -0.34σ, respectively.

Fig. 7.— Histograms of simulated flux estimates for the 5-parameter model of Figure 5, with count rates adjusted to give $\approx$ 100, $\approx$ 1, and $\approx$ 0.01 counts per bin. The means and widths of the histograms agree with the theory independent of count rate. For the low count rate case, about 70% of the scans have no counts at all; the shape of the distribution is discussed in the text.

Fig. 8.— Typical histogram of *HEAO 3* single-scan flux estimates, standardized as described in the text, for a source with no detectable flux.



LOG (RATE [COUNTS s-1 keV-1])
1
0
-1
-2
-3
-4
-5
-6
1
2
3
4
LOG (ENERGY [keV])
Background
CRAB
1 SIGMA
POISSON NOISE

Fig. 1

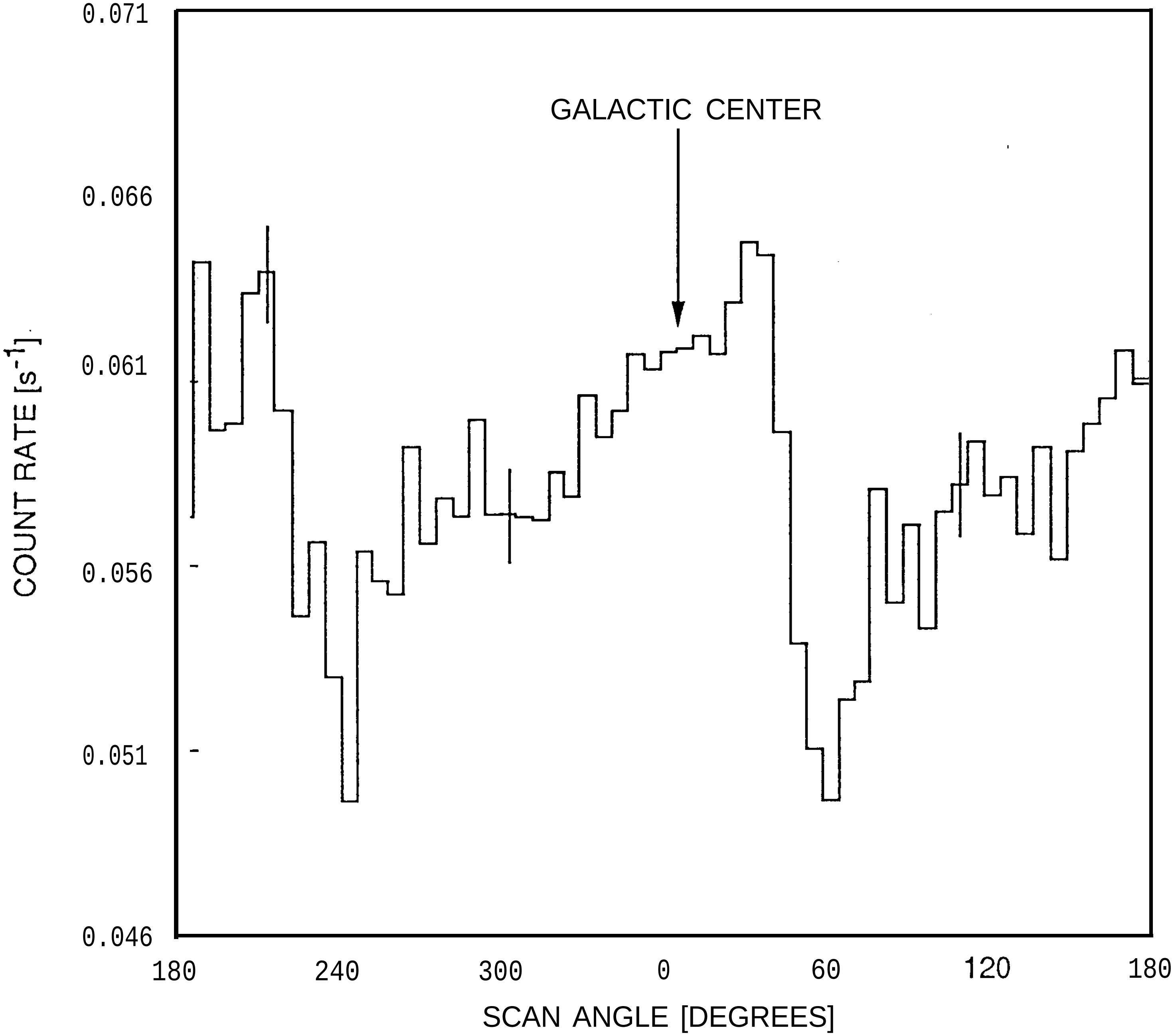

GALACTIC CENTER
COUNT RATE [s^-1]
0.071
0.066
0.061
0.056
0.051
0.046
180
240
300
0
60
120
180
SCAN ANGLE [DEGREES]

Fig. 2

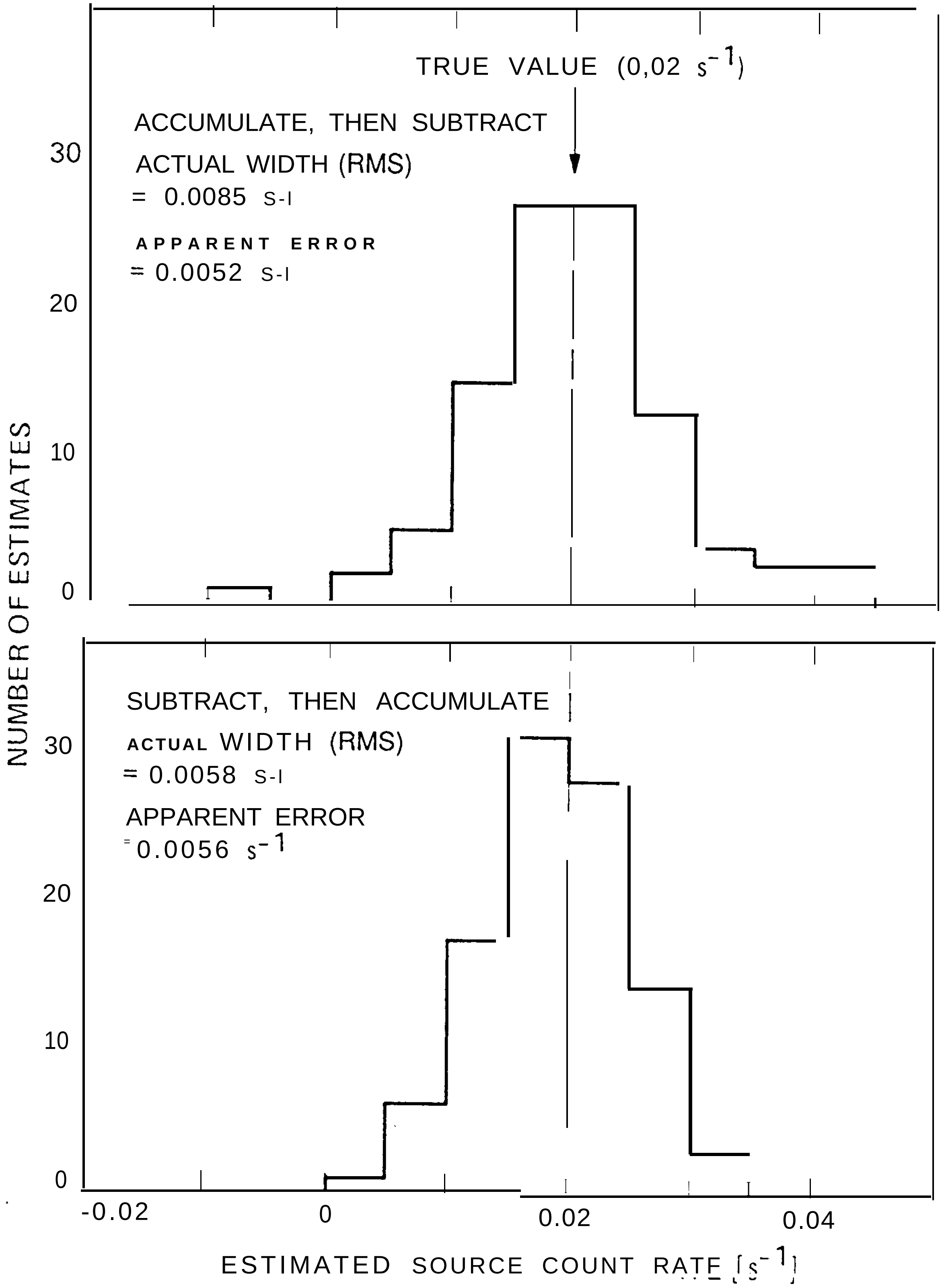

TRUE VALUE (0,02 s-1)
ACCUMULATE, THEN SUBTRACT
ACTUAL WIDTH (RMS)
= 0.0085 s-1
APPARENT ERROR
= 0.0052 s-1
SUBTRACT, THEN ACCUMULATE
ACTUAL WIDTH (RMS)
= 0.0058 s-1
APPARENT ERROR
= 0.0056 s-1
NUMBER OF ESTIMATES
0
10
20
30
-0.02
0
0.02
0.04
ESTIMATED SOURCE COUNT RATE [s-1]

Fig. 3

ThuApr2901 21:591993 Wheaton et. al. Fig 4

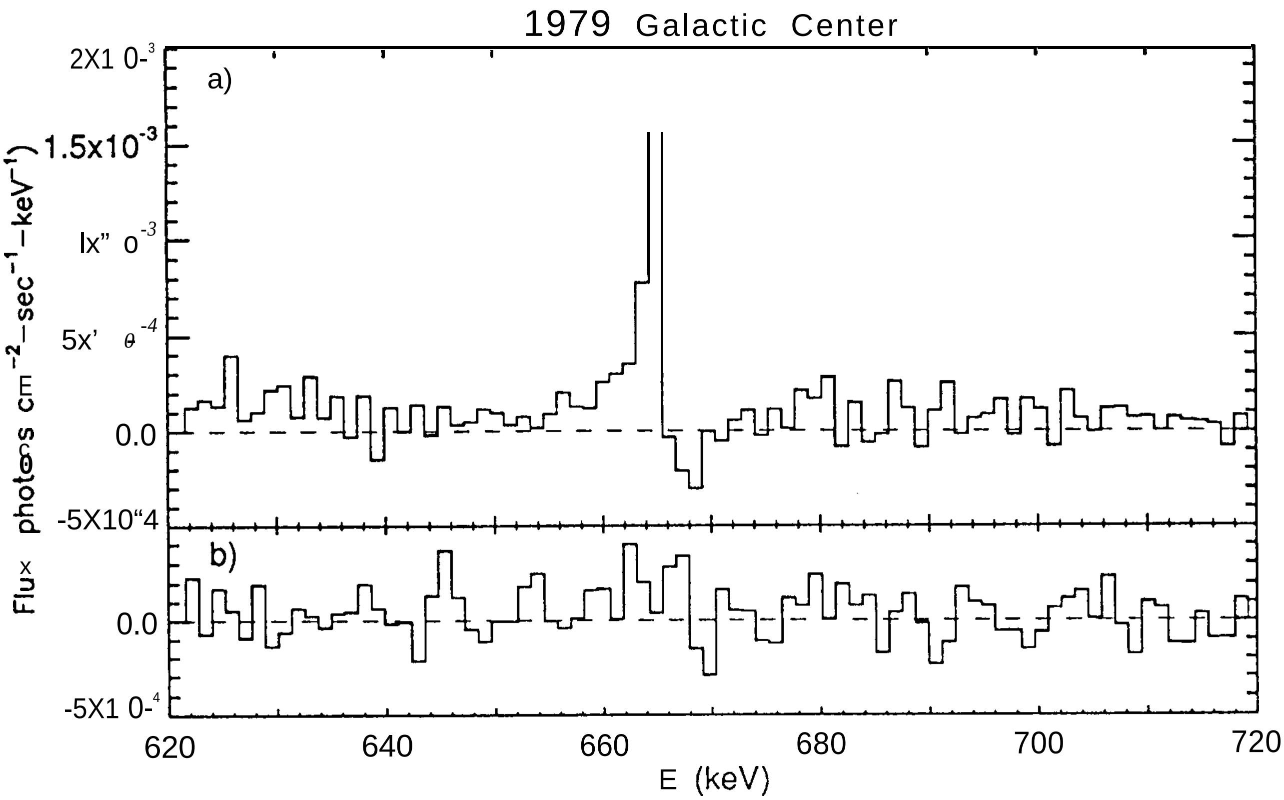


Figure 4

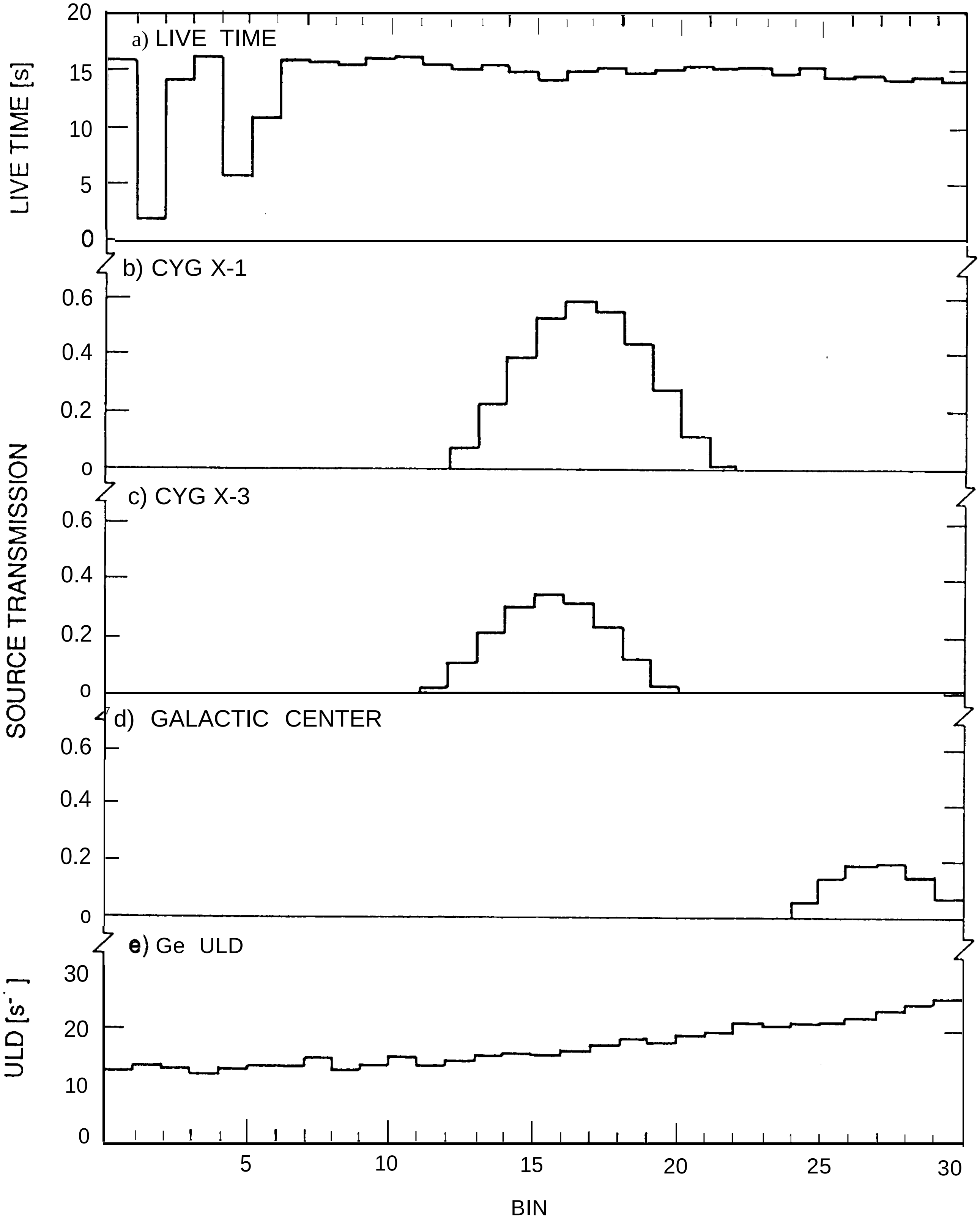
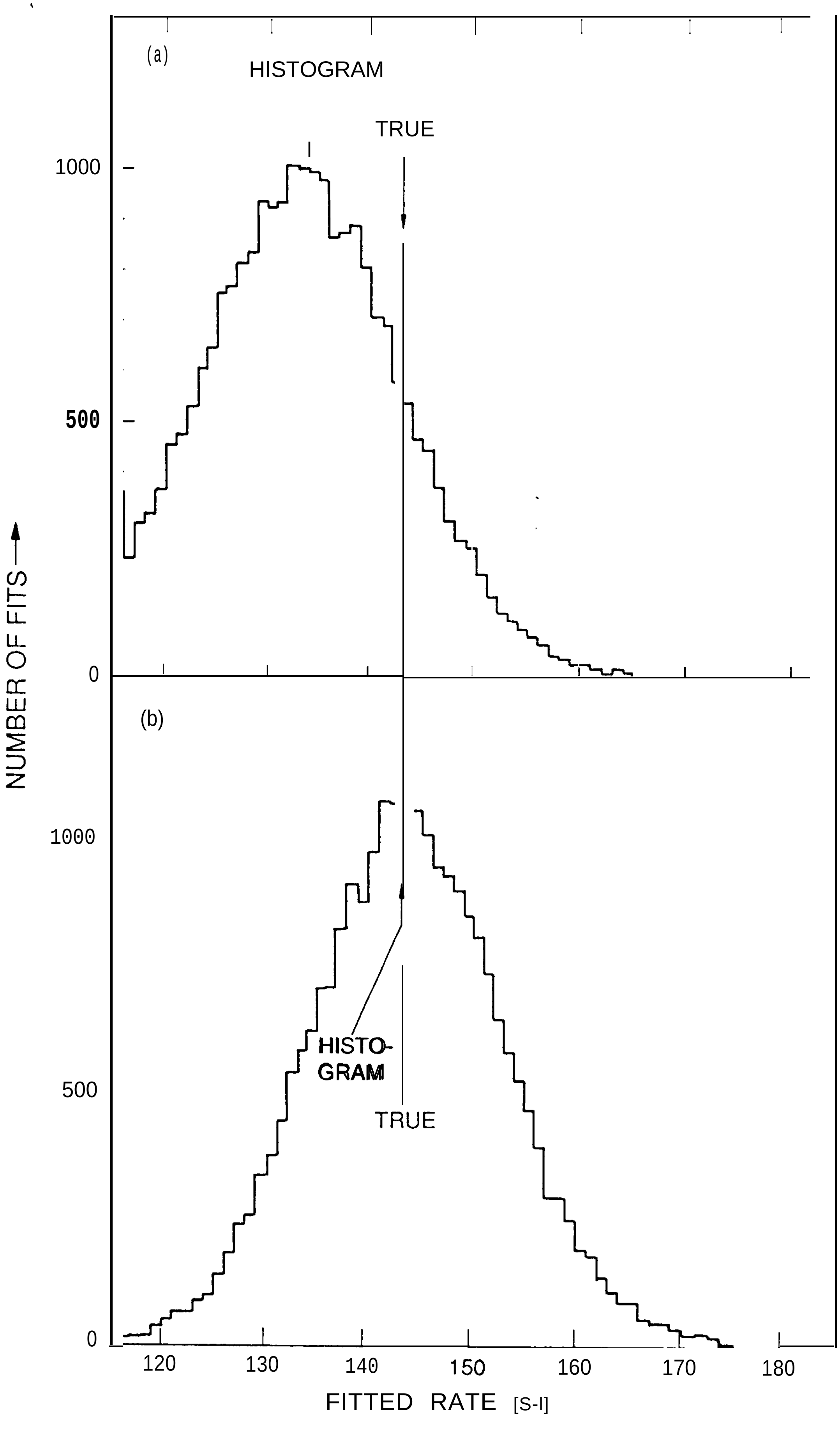
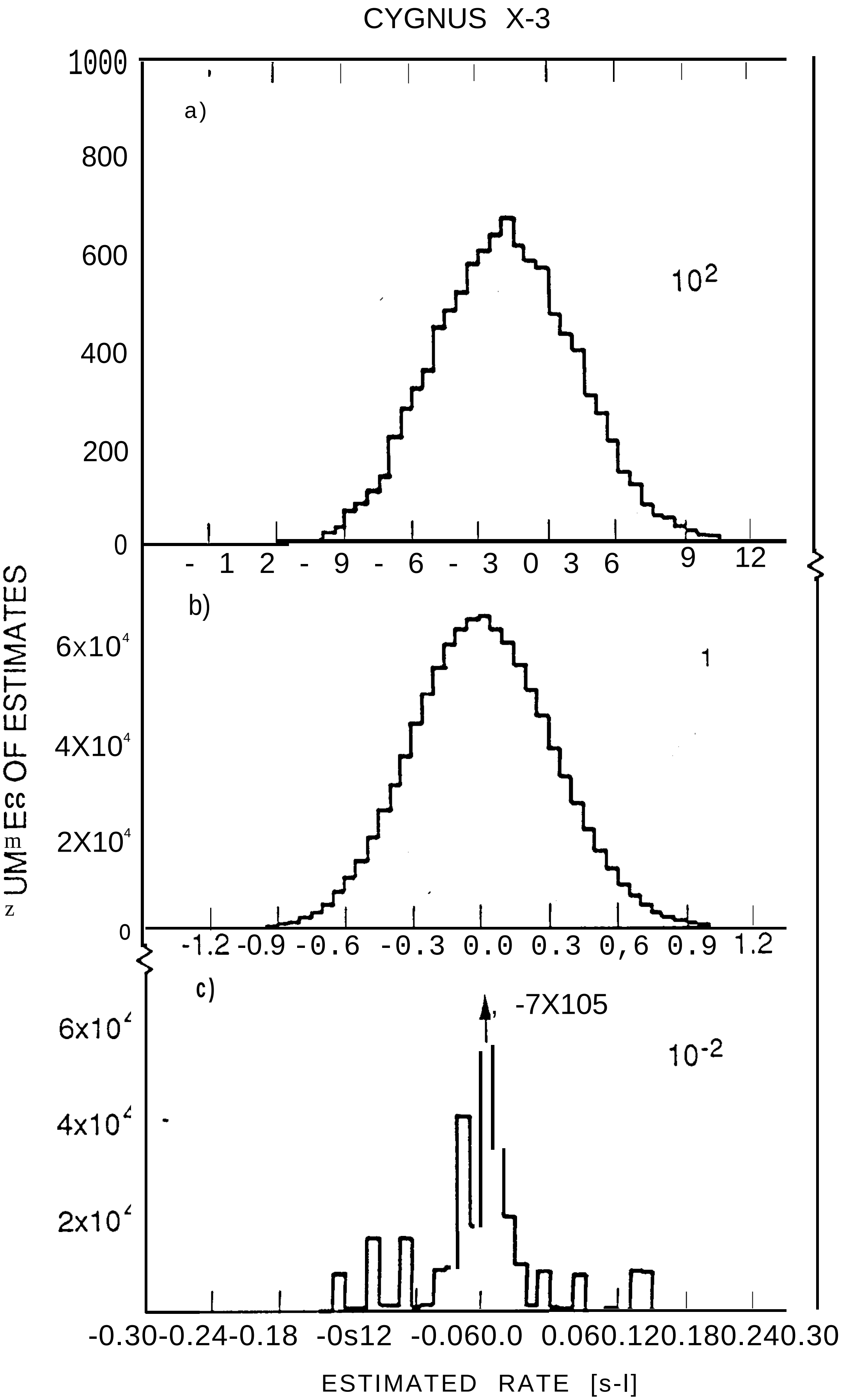

a) LIVE TIME
LIVE TIME [s]
b) CYG X-1
c) CYG X-3
d) GALACTIC CENTER
SOURCE TRANSMISSION
e) Ge ULD
ULD [s⁻¹]
BIN
0
5
10
15
20
25
30
0.2
0.4
0.6

Fig. 5



(a)
HISTOGRAM
TRUE
1000
500
0
(b)
1000
HISTO-
GRAM
500
TRUE
0
120
130
140
150
160
170
180
NUMBER OF FITS
FITTED RATE [S-I]

Fig. 6



CYGNUS X-3
a)
1000
800
600
400
200
0
10²
- 1 2 - 9 - 6 - 3 0 3 6 9 12
b)
6x10⁴
4X10⁴
2X10⁴
0
1
-1.2 -0.9 -0.6 -0.3 0.0 0.3 0,6 0.9 1.2
c)
-7X105
6x10⁴
4x10⁴
2x10⁴
10⁻²
-0.30-0.24-0.18 -0s12 -0.060.0 0.060.120.180.240.30
ESTIMATED RATE [s-l]
UMBECC OF ESTIMATES

Fig 7

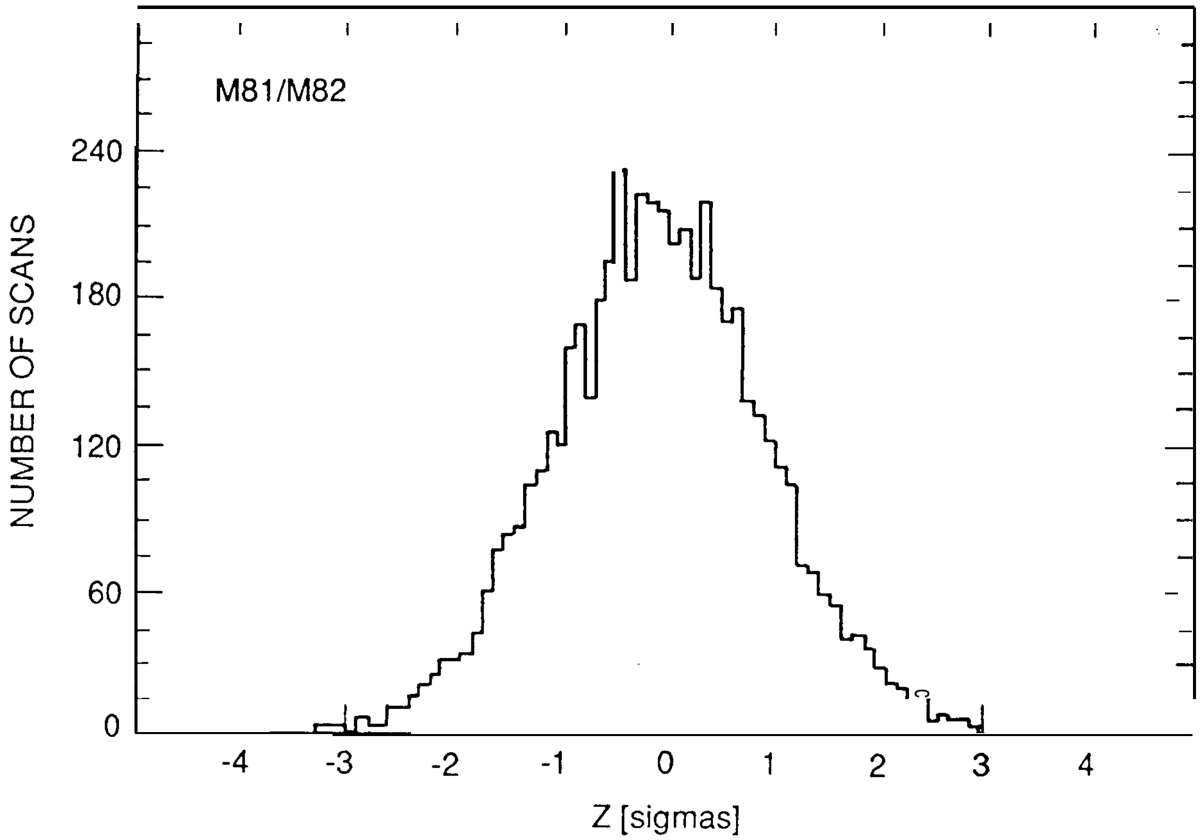


Fig. 8